Faculty of Health Sciences

Student evaluation practice

A qualitative study on how student evaluation of teaching, courses and programmes are carried out and used

Iris Helene Borch

A dissertation for the degree of Philosophiae Doctor   March 2021

UiT The Arctic University of Norway

Student evaluation practice: A qualitative study on how student evaluation of teaching, courses and programmes are carried out and used

Iris Helene Borch

Faculty of Health Sciences
Centre for Teaching Learning and Technology
UiT The Arctic University of Norway
Tromsø 2021

Table of Contents

Acknowledgements

Almost four years of the PhD period has come to an end and it is time to acknowledge the people who have contributed to the project in different ways. It has been a wonderful journey thanks to all the supportive, knowledgeable and inspiring people I am surrounded with.

First and foremost, I would like to express my deepest gratitude to the informants in this study. I am grateful for your contribution, for sharing your perspectives and experiences with student evaluation practice.

I would like to thank UiT the Arctic University of Norway for funding the project and Result, the Centre for teaching, learning and technology for providing a good working environment. In particular, thank you to Ådne Danielsen for valuable discussions about student evaluation practice, and help and support with the theoretical parts of the second paper. I also want to thank graphic designer Mark Stenersen, at Result for creating the visual figures in the thesis. I would like to thank my former colleagues and fellow PhD students Siri, Jens and Pål Anders who included me in a PhD fellowship at "the 5th floor", for our discussions and your tips for where to seek advice in the journey of becoming a scholar. A special thank you to Anka, my current PhD fellow and friend. You have contributed positively to my working environment, on campus but also during our writing retreats off campus, thank you for our many discussions, walks, coffee breaks, dinners and lunches. I am going to miss you a lot. Thank you to all the members of the newly established research group at Result for inspiring discussions and feedback on my project. I would also like to thank my native English speaking colleagues Michael and Ellen for contributing with clarifications and suggestions for how to improve the language. Thank you to library staff for EndNote support and administrative employees who have helped me access documents and provided contextual background information during the project period.

A warm thank you to Tine Prøitz who conducted the midterm and the end of term evaluation of the project. Your constructive feedback helped me see my work from new perspectives that truly also improved the dissemination of the findings and the quality of the extended abstract.

Ragnhild Sandvoll and Torsten Risør, I am grateful for everything you have contributed with as my supervisors. You have supported and challenged me, listened and given valuable feedback, asked critical questions and helped me move wisely forward and becoming an

independent scholar. Ragnhild, my supervisor, thank you for your enthusiasm, quick responses to my many questions, positivity, extraordinary care and for being a role model. Torsten, my co-supervisor, thank you for providing new perspectives that helped me broaden my understanding, for well-formulated re-writings, philosophical reflections that sometimes made it necessary to pause and rethink. Together, I think we have been a great team!

I would also like to thank the Centre for Health Education Scholarship (CHES) at University of British Columbia, Vancouver, Canada for generously providing an inspiring and developing working environment during my stay as a visiting scholar at CHES in 2017/18. Thank you to the National Research Council of Norway who funded the stay abroad for me and my family.

A warm thank you to my former colleagues at the Faculty of Health Sciences who motivated me to apply for the PhD project, particularly Anita Iversen who inspired me to engage in the field of academic development long before I became a PhD student. Thanks to Anita and Marit Allern for reading an early draft of the project description and for providing valuable feedback. Thank you to former colleagues at the Bachelor in Physiotherapy programme for your dedication to educational quality and student learning. Your enthusiasm was "contagious" and an inspiration for this project. Grete Barlindhaug deserves an extra thank you for your contribution as an observer during the focus group discussions.

I am grateful for all my dear friends and family who have been supportive and patient, who have helped me balance work life and my social life in a good way.

A warm thank you to my parents for always being there for me and my family, thank you for your eternal support, love and care. I would also like to thank my parents in law for your support and for providing a perfect writing space in Ravik where many paragraphs of the thesis are written.

Last but not least, thank you to my dear husband Fredrik and our beloved children Olav, Simon and Noah. Your patience, encouragement, support and love mean everything to me. I truly look forward to spending more time together with you, to be more present in our everyday life and for creating new memorable moments with my favorites.

Iris H. Borch, Tromsø, March 2021

Abbreviations

AD	Academic Developer
CA	Constructive Alignment
CAQDAS	Computer-Assisted Qualitative Data Software
ECB	Evaluation Capacity Building
ENQA	European Agency for Quality Assurance
ERIC	Educational Research Information Centre
ESG	The Standards and Guidelines for Quality Assurance in the European Higher Education Area
NFR	Norges forskningsråd (Norwegian Research Council)
NOKUT	Nasjonalt Organ for Kvalitet i Utdanningen (Norwegian Agency for Quality Assurance in Education)
NPM	New Public Management
QA	Quality Assurance
QAS	Quality Assurance System
QE	Quality Enhancement
SET	Student Evaluation of Teaching

List of papers

Paper 1:

Borch, I., Sandvoll, R., & Risør, T. (2020). Discrepancies in Purposes of Student Course Evaluations: What Does It Mean to Be "Satisfied"? *Educational Assessment, Evaluation and Accountability*, *32*(1), 83-102. https://doi.org/10.1007/s11092-020-09315-x

Paper 2:

Borch, I. (2020). Lost in Translation: From the University's Quality Assurance System to Student Evaluation Practice. *Nordic Journal of Studies in Educational Policy*, *6*(3), 231–244. https://doi.org/10.1080/20020317.2020.1818447

Paper 3:

Borch, I., Sandvoll, R., & Risør, T. (2021). Student Course Evaluation Documents: Constituting Evaluation Practice. *Assessment & Evaluation in Higher Education*. https://doi.org/10.1080/02602938.2021.1899130

List of figures

List of tables

Abstract

Background/aims: This PhD project explores how internal student evaluation of teaching, courses and programmes at eight health profession education programmes are carried out and used. Student evaluation is a mandatory part of local quality assurance systems in Norway and aims to be used in educational quality assurance and enhancement. At the university of this study are these evaluations also considered to be part of students' learning processes and student empowerment. The study investigates pedagogical and organisational dimensions with evaluation practice, including how different actors interact in evaluation processes and how the interplay between policy and practice are balanced.

Methods: The study has an exploratory qualitative research design and comprises three different research methods: semi-structured interviews with academics, focus group interviews with students and document analysis of internal evaluation documents.

Results: The study consists of three papers, each exploring different aspects with student evaluation practice based upon different empirical data and analytical perspectives. Together the papers revealed that there are both organisational and pedagogical dimensions with evaluation practice that seems to affect how evaluation is carried out and used. The study discovered a conspicuous gap between intended use and practice articulated in the local quality assurance system, and use described by the academics and students. The study shows aspects with evaluation practice that seem to limit use of evaluation for educational quality purposes. Some of these limiting aspects were categorised as organisational dimensions. Examples are limited communication about student evaluation within the programmes and across organisational levels, low sense of ownership to evaluation guidelines among academics and lack of student perspective in educational quality reports documenting student evaluation practice. The guidelines and evaluation system are developed by administrative staff on behalf of the university management with an expectation that academics will follow these. The academics were left to themselves when carrying out and following up on evaluations and expressed a need for more support and knowledge about evaluation. Some of these organisational dimensions also affect the opportunity to carry out evaluation practices that are part of students' learning processes, e.g., too little time to follow up on evaluation results and establish evaluation practices that invite students to provide feedback about their learning processes. In the study aspects with the methods themselves are categorised as pedagogical

dimensions affecting how evaluation is used. The study reveals that the evaluation questions, particularly in surveys were teacher- and teaching-focused and satisfaction based rather than student- and learning-focused. Consequently, students' responses from surveys are better suited for quality assurance than quality enhancement. Dialogue-based evaluation methods have a more open format and invite students to reflect upon their learning processes. Students consider these dialogues valuable for their professional development and the academics expressed that they used students' feedback for adjustments of the teaching approaches.

Discussion: Based upon the findings and an understanding of student evaluation as processes that can be used to promote educational development, I have developed a guiding framework for universities that want to strengthen the learning focus in student evaluation. This framework is also used to illustrate today's practice. The framework illustrates how internal student evaluations are complex processes dependent on interaction between administrative staff, academics and students. I discuss how pedagogical and organisational dimensions with evaluation practice can be strengthened if student evaluation shall be embedded in a student-learning-centred evaluation practice.

Contribution: The interaction between different actors across organisational levels and between policy and practice should not be underestimated if evaluation is going to be part students' learning process, student empowerment, and also be used in both quality assurance and enhancement. Today's student evaluation practice seems to be carried out in ways addressing and ensuring educational quality more than it promotes quality development, individual and organisational learning.

1.0. Introduction

Eight years ago, two months after I was employed at the university as a head of studies of a bachelor programme, when I was planning a diploma ceremony, an administrative staff colleague sent me a friendly reminder that I had to conduct a programme evaluation with the students. I asked my colleague whether the university or faculty used standardised surveys for programme evaluations but learned that each programme used their own. I asked other colleagues and the teachers at the programme if they had examples of surveys they had conducted earlier and which topics or questions they believed we should include. Based upon their feedback and examples of evaluation surveys, I created a survey that was sent to the students before the graduation ceremony, and I eagerly awaited for response. Patiently. The response was disappointingly low. The feedback from the few students who responded gave very little information and raised new questions about the students' perceptions of the programme and their learning outcomes. These notions sparked an interest in me to explore student evaluation and I started to reflect upon what I could use this evaluation data for—about purposes and uses of student evaluation with a motivation to improve my evaluation approaches. This was my first experience with student evaluation as an employee at the university and the beginning of my engagement and a long endeavour exploring the complex phenomena of this PhD project: student evaluation and its uses.

1.1 Background

In contemporary society we are evaluated and asked to give feedback in almost all arenas of our lives. Both informal and formal evaluations are increasing in numbers. The term "evaluation" is used to describe many different methods, strategies, processes, policies and activities that take place on individual and organisational levels in different contexts. Consequently, it is a challenging concept to define (Schwandt, 2009). Evaluation has been described as "assisted sensemaking" (Mark, 2009, p. 55) that with the help of "'artificially' constructed methods and procedures, helps construct data which indicate whether particular activities are good or good enough, whatever that means in particular contexts" (Dahler-Larsen, 2005, p. 615). This thesis explores student evaluation of teaching, courses and programmes and how these evaluations are used at a Norwegian university, from the perspectives of academics and students. Thus, it investigates what evaluation means to actors involved in internal

evaluation. The context is UiT the Arctic University of Norway. I start with a brief introduction about evaluation to get a better understanding of the complexity and multiple functions the term evaluation comprises, before delving into student evaluation. I am therefore including different ways of defining evaluation in this introduction section. It may be helpful as a starting point to distinguish between an everyday use of the concept evaluation and a more formal evaluation. The former often refers to human sensemaking and evaluative judgement (Mark, 2009) and the latter "uses *formal* methodologies to *provide useful empirical evidence* about *public entities* (such as programs, products, performance) to provide empirical evidence in *decision-making contexts* (…) (Trochim, 1998, p. 248; emphasis in original). Scholars and evaluators have developed many other definitions of evaluation that emphasise different aspects of evaluation. One of the most cited definitions of evaluation is by Michael Scriven (1991, p. 139): "Evaluation refers to the process of determining the merit, worth or value of something, or the product of that process". This definition points to the goal of evaluation that can be described as "to consider value in a systematic way" (Vo & Alkin, 2017, p. 10).

I understand evaluation as a social phenomenon and as a practice wherein actors and contexts involved in the evaluation processes are important. Hence, evaluation is emphasised as processes and not as single instruments, models or standards. Evaluation approaches that build on these characteristics of evaluation were labelled by Guba and Lincoln (1989) as fourth-generation evaluation; these evaluation approaches are grounded in constructivism. This understanding of evaluation represented a counterpart to earlier generations of evaluation that were based upon extensive use of testing and measurement (first generation), use of objects and tests (second generation), judgement and decision-based evaluation models (third generation) (Guba & Lincoln, 1989). Evaluation outcomes are, in fourth-generation evaluation, "meaningful constructions that individual actors form to 'make sense' of the situations in which they find themselves" (1989, p. 8), and valuing is regarded as "an intrinsic part of the evaluation processes providing the basis for attributed meaning" (1989, p. 109).

Evaluation inquiries are strongly related to the purposes of evaluation itself. Most evaluators will agree that we do evaluations to learn something about what is being evaluated that in turn can help the programme, product or policy to improve. Evaluation scholar Michael Quinn Patton, the founder of Utilization-Focused Evaluation[1] (UFE), states that evaluation inquiries

---

[1] Utilization-Focused Evaluation is presented in Chapter 3.2 and Paper 1.

depend on their purposes (2008). From this point of view, internal educational evaluations must clearly be different from for example, industrial product evaluations because of their different purposes. Defining educational evaluation in a precise way can be nearly impossible (Kellaghan & Stufflebeam, 2012; Schwandt, 2009). One reason it is hard to define educational evaluation is that evaluation has many potential users: educational leaders, administrative staff, educational politicians, students, academics and university management. Another reason it is a challenging term to define is that educational evaluation comprises many formats: student evaluation of teaching, teacher evaluation, course evaluation, programme evaluations etc. Nonetheless, they have a common purpose: "The goal of all educational evaluation is to enable programs and policies to improve student learning" (Ryan & Cousins, 2009, p. ix). As this is a thesis that explores the phenomenon student evaluation practice, specifically the practice of student course and programme evaluation and evaluation of teaching, I will in the remainder of the thesis focus on student evaluation and frequently simply use the term "evaluation". It is recognised that evaluation of higher education teaching and programmes needs to draw on a number of sources rather than relying purely on student feedback (Berk et al., 2005; Cathcart et al., 2014). I acknowledge that student evaluation is just one of many concepts that aim to judge, describe, assure and improve educational quality.

Evaluation has been institutionalised as a phenomenon that many people take for granted will occur in modern organisations (Dahler-Larsen, 2011, p. 2). This is also the case in higher education. The type of educational evaluation explored in this PhD project is internal student evaluation of teaching, courses and programmes—evaluations that are initiated, carried out and followed up at the university. These are in other words micro-evaluations, but as these evaluations are mandated and essential to national educational policy some may also regard them as meso-evaluations. I rely on a definition by Hanne Foss Hansen (2009, pp. 72-73) about evaluation at different organisational levels; she defines micro-evaluations as local initiated evaluations that are "embedded in learning environment and includes both dialogue between students and teachers and more systematically procedures for testing and assessing students and sometimes other types of stakeholders assessing teachers and individual programmes. Meso-evaluation is defined as evaluation institutionalized as an element in national educational policy".

It is common to divide evaluation into formative and summative. Scriven introduced as early as the 1960s the distinction between formative and summative evaluations (Scriven, 1967). The

distinction between formative and summative evaluation remains in higher education today and can be regarded as established terms in the sector (Patton, 2008; Scriven, 1991, 1996). However, the widespread nature of this distinction has been interpreted slightly differently from the original definition (Patton, 2008). Therefore, I consider it useful to cite Scriven (1991, pp. 62, 150) and his definitions;

> "Formative evaluation is conducted during the development or improvement of a program or product (or person etc.). It is an evaluation which is conducted for the in-house staff of the program and normally remains in-house; but it may be done by an internal or external evaluator or (preferable) a combination". (1991, p. 62)
> "Summative evaluation of a program (etc.) is conducted *after* the completion and for the benefit of some *external* audience or decision-maker (…), though it may be *done* by either internal or external evaluator, or a mixture." (1991, p.150)

In formative evaluations, the purposes are to conduct information that can be used for learning and improvement. This differs from summative evaluations that aim to judge effectiveness and are often used in decision making (Patton, 2015; Vo & Alkin, 2017).

Most of research on student evaluation is conducted in the USA, Australia and the UK, contexts that are quite different from those in Scandinavian countries. Both the contexts and the ways student evaluation are used differ between America, Australia and the UK, and Scandinavia. The education systems in the former contexts request high tuition fees from enrolled students, particularly because many of these are private institutions, whereas the majority of the Scandinavian education system is governed and strongly regulated by the State. These differences between countries with high public regulation and countries with lower public regulation have also affected how quality assurance has been played out in the different contexts (Elken & Stensaker, 2020b; Stensaker & Harvey, 2011). This also affects the position student evaluations have in society and how these evaluations are used. Some major differences between how student evaluation is used are: by students selecting the institution where they want to study and by universities for administrative purposes, national rankings and the level of standardisation. Evaluation plays a more important role in administrative processes like hiring, tenure and salary and national rankings, and evaluation practice is more standardised in the USA and UK than in Scandinavian countries.

Although the contexts where the majority of evaluation research is conducted differ from Scandinavia, the major principles of student evaluation are the same in most contemporary education systems. Evaluation researcher Stephen Darwin (2016, p. ix) says:

"At its essence, student evaluation necessitates a judgment being exercised from a particular viewpoint (the subject) on an identified and bounded entity (the object). Conventional quantitative forms of student evaluation invite the judgment of individual students to be exercised on the value of teachers, teaching approaches and courses at the end of the semesters. The criteria for such judgments are inherently subjective, but its outcomes are objectively framed in numeric rating scales that form the basis of student feedback reports. The explicit intention of these student feedback reports is to inform future academic decision-making."

Student evaluation (of teaching, courses and programmes) has manifested its position in Norwegian higher education by legal regulations (Universitets- og Høgskoleloven, 2005). Student evaluation is described as essential to quality assurance of higher education in a national white paper, "the Quality reform", from 2001 (Kirke- utdannings- og forskningsdepartementet, 2001) and became mandatory by law in 2002 as part of the institutions' quality assurance systems (QAS). The intention with the implementation of local quality assurance systems was to assure a continuous improvement of educational quality, wherein student evaluation is essential (Kunnskapsdepartementet, 2007, 2017). From a Norwegian educational policy perspective, the intended purposes of the student evaluation data set by The Ministry of Education and the National Quality Assurance Agency (NOKUT) are educational quality enhancement (QE) and quality assurance (QA). However, the Ministry acknowledges that these are not used as intended for educational improvement (Kunnskapsdepartementet, 2017). The Ministry points at academics' scepticism towards relying too much on students' feedback when judging educational quality but in the white paper they do not provide any explanations or refer to research that explain why this might be the case or how to improve educational evaluation practice. Academics' scepticism towards student evaluation has also been expressed in chronicles published in the national university press, Khrono (Larsen, 2020), as a response to a national external student experience questionnaire, Studiebarometeret[2] (NOKUT). Central to this debate was the discussion of what student evaluation data can be used for, what kind of data Studiebarometeret collects and how different actors involved in student evaluation processes view evaluation differently. Like the statement in the white paper, this debate also lacked empirical references from Norwegian higher education. This study aims to contribute knowledge about student evaluation practice, including

---

[2] Studiebarometeret is an annual national student experience questionnaire developed and administered by the national agency of higher education, NOKUT, for an English presentation of Studiebarometeret: https://studiebarometeret.no/en/. The debate about Studiebarometeret in Khrono lead to a webinar run by NOKUT in November 2020. Academics who had written chronicles in Khrono were invited to have presentations at the webinar as an introduction to a discussion session. More about the webinar in this article in Khrono (Larsen, 2020): https://khrono.no/kritikere-krever-endringer-i-studiebarometeret/528746.

reasons why academics are sceptical of relying too much on student evaluation when judging educational quality. Further, this study adds empirical knowledge to an ongoing national and international debate about student evaluation balancing between control and accountability on one hand and enhancement and learning on the other.

Low use of evaluation data is not exclusive to Norwegian higher education. Despite ambitions to use student evaluations to improve student learning and teaching, the actual rates of use for this purpose are found to be low (Beran & Rokosh, 2009; Beran & Violato, 2005; Darwin, 2017; Kember et al., 2002). Notwithstanding this identified gap between a belief and trust in student evaluation as a policy strategy for improved educational quality on one hand and low use on the other, student evaluation has manifested its position in higher education. Norwegian higher education institutions are within the confines of the law, yearly conducting feedback from students in high numbers with a plethora of evaluation approaches. Students are invited to provide their feedback about their education using a variety of methods. Nevertheless, there exists little knowledge about student evaluation from the students' perspective (Darwin, 2016). There exists more knowledge about how teachers perceive evaluation than students. Different stakeholders like academics, students and administrative staff have divergent understandings of what constitutes educational quality (Dicker et al., 2019; Oliveira et al., 2012) and good teaching (Nasser-Abu Alhija, 2017; Prosser et al., 2003). Therefore, is it important to balance different perspectives when evaluating educational quality and to explore how student evaluation practice is understood by the key stakeholders who are providing feedback about educational quality. With a social constructivist approach underscoring how knowledge is constructed by actors, I aim to explore student evaluation from different perspectives in order to get a better understanding of student evaluation practice. My understanding of practice of evaluation relates to Saunders (2011, p. 2) who suggests it can "usefully be conceptualised as sets of clusters of behaviours forming ways of thinking and doing associated with undertaking evaluative activity, this includes the rooted identities and patterns of behaviours that characterise shape and constrain understanding of evaluative practice".

Review studies and my own literature reviews show that prior research has been dominated by quantitative studies exploring aspects of student evaluation methods, particularly aspects like bias, validity and reliability with evaluation surveys (Abrami et al., 2007; Alderman et al., 2012; Richardson, 2005) and fewer qualitative studies that explore different aspects of evaluation practice from stakeholders' perspectives. Furthermore, researchers have investigated aspects of

the tools like the validity, reliability, response rates etc., rather than doing research on how to best use evaluation (Penny, 2003). In short, researchers have explored fragmented parts of evaluation rather than exploring evaluation practice and how evaluation is carried out, i.e., how actors interact with each other at different organisational levels and the balance between policies and practice.

Based upon literature reviews on student evaluation presented in section 2.0 and the abovementioned contextual status, I aspire to explore student evaluation *practice* at the university and incorporate both organisational and pedagogical dimensions of evaluation. In this thesis I explore student evaluation and its uses from different perspectives at UiT, the Arctic University of Norway, with the intention that this study can contribute knowledge that has implications for future evaluation practice. The aim of the thesis and research questions are presented below.

1.2 Research questions

The aim of this thesis is to explore internal student evaluation practice and how evaluation is carried out and used at the university. The focus is on students' and academics'[3] perspectives on student evaluation practices, as well as how evaluations feature in internal documents. I aspire to contribute knowledge about informal and formal, pedagogical and organisational dimensions of evaluation practice from different actors' perspectives. The overarching research question for the PhD project is:

How are student evaluations carried out and used at UiT the Arctic University of Norway?

Based on this overarching research question, the following sub-questions were posed in the three papers:

- How do different evaluation methods, such as survey and dialogue-based evaluation, invite students to provide feedback about aspects relevant to their learning processes? *(Paper one)*
- How is evaluation contextualised and translated locally at the university? *(Paper two)*

---

[3] Academics in this study are academic leaders on the programme level, responsible for courses and/or programmes. The terms academics, teaching academics and leaders are used interchangeably in different contexts and papers.

- How are student evaluations documented and reported and how can internal evaluation documents contribute to the constitution of evaluation practice? *(Paper three)*

Each of the papers explores how evaluations are carried out and the uses of internal evaluation in relation to different dimensions/aspects of evaluation practice, respectively. They complement each other in that they investigate different parts of evaluation practice from different perspectives, using different methods and analytical perspectives to achieve a comprehensive answer to the overarching research question. The three papers together contribute knowledge about pedagogical and organisational dimensions of evaluation practice and explore use in relation to evaluation methods (paper one), use in relation to how academics translate evaluation into practice (paper two) and use in relation to documentation routines (paper three). There are different approaches to studying use. Use in this study is not about the researcher observing use of student evaluation but it is about uses, influence or outcomes of evaluation from the perspectives of students and academics, described in interviews, as well as how they appear in documents describing evaluation. Consequently, use in this study refers to the kind of use that students and academics identify and elaborate on in the interviews, particularly on the programme level, as well as use identified by me as a researcher by analysing the interviews and documents from institutional, faculty, department and programme level.

Definitions of evaluation use have changed and been debated by evaluation scholars over decades; these changes, different understandings and types of evaluation use will be elaborated on in section 3.2. To start with, I am providing a definition of evaluation use by Vo and Alkin (2017, p. 265) who state: "evaluation use refers to the way in which the evaluation process and the information obtained from an evaluation impacts the program that is being evaluated". The understanding of evaluation use in this thesis also incorporates the term "evaluation *influence*" provided by Kirkhart (2000). She proposed the term to expand the possibilities of evaluation use beyond direct use of evaluation data and processes, and to include "*in*direct, *in*tangible influence that evaluation studies can have on individuals, programs, communities, and systems" (Alkin & King, 2017, p. 443). Kirkhart (2000, p. 5) proposed the term "evaluation influence" in addition to better foster an "inclusive understanding of the impact of evaluations". Uses explored in this thesis are unintended and intended uses described by the informants in the interviews and by me through document analysis. The informants describe use with their everyday language and how they experience use of student evaluation. However, I apply terms

from evaluation use literature as an analytical perspective to describe how evaluations are carried out and used. The explored uses happen at different levels in the organisation and are described by different actors within the university[4].

1.3 Educational quality

Veronica Bamber and Sally Anderson (2012, p. 6) state that: "The story about evaluation in universities is the story of quality". Actors in internal evaluation in higher education would probably easily agree with them because evaluation is regarded as essential to judging, enhancing and assuring educational quality. As evaluation and quality strongly relate to each other and student evaluation is central to quality work, quality assurance and quality enhancement, I am in this subsection providing short definitions of these terms.

Educational quality

Since the late 1980s, educational quality has been an overall aim for contemporary universities (Bleiklie, 1998). How to define and judge educational quality in higher education has since then[5] been debated and is still debated without a common agreement having been established (Harvey & Green, 1993; Harvey & Williams, 2010; Schindler et al., 2015; Wittek & Kvernbekk, 2011). There is, however, a consensus that different actors in higher education like students, teachers, management and policy makers attribute different meanings to what constitutes good educational quality (Westerheijden, Stensaker, et al., 2007). Harvey and Green (1993) suggested that 'Quality' is a philosophical concept that is understood differently by different people. They created one of the most frequently cited definitions of quality and grouped the different understandings into five categories: exceptional, perfection, fitness for purpose (purposeful), value for money (accountable) and transformation (transformative). Newer understandings of educational quality often incorporate "a stakeholder-driven" definition of quality (Schindler et al., 2015).

The understanding of educational quality that I rely on in this thesis is the one used in the local quality assurance system. This definition is strongly related to definitions in Norwegian higher

---

[4] I would like to note that although some of the included reports in this study are written as part of a documentation requirement and address an audience outside the organisation, I am studying evaluation practice at the university.

[5] Christopher Ball published an essay (1985) entitled "What the hell is quality?", this essay is by many considered to have sparked the debate about educational quality in the 1980s.

education policy documents and definitions used by the Norwegian agency for quality assurance in education (NOKUT)[6]. It sprung from a definition introduced in an Official Norwegian Report (NOU) in 2000 by an advisory committee to the Norwegian Ministry of education (Mjøs & Utvalget for høgre utdanning, 2000), and the definition has since then been debated and complemented. Newer national policy documents provided by the Ministry of Education (Kunnskapsdepartementet, 2017) and NOKUT (2016) provide recommendations as to how to understand educational quality rather than definitions. The understanding of educational quality stated by NOKUT today comprises an understanding of educational quality as a process with partly overlapping subcategories. Seven subcategories that build upon the description provided by NOKUT (2016) and the Ministry of education (Kunnskapsdepartementet, 2017) are articulated in the local quality assurance system at UiT (2012): Programme Quality (1), Teaching Quality (2), Relevance Quality (3), Framework quality (4), Management quality (5), Entrance Quality (6) and Result Quality (7).

Although it is not explicitly expressed in QAS which quality aspects students are expected to give their feedback about, it is likely from my interpretation and understanding of the system that student evaluation aims to get feedback about the perceived quality of a study programme (1), teaching (2) and its relevance (3) and frameworks (4)—in other words, four of the seven categories. I will cite the definition of these categories provided in QAS [7] (UiT, 2012, p. 2):

(1) Programme quality covers the programme's academic content and the organisation of the various components that form part of the programme.

(2) Teaching quality covers the academic and pedagogical level and the implementation of the educational activities and academic supervision.

(3) Relevance quality covers the relevance of the education in relation to society and the employment arena's competence requirements, as well as long-term added value that the courses and programmes contribute to society, culture and the individual student.

(4) Framework quality covers the university's work with the respect to the physical, psychological and organisational learning environment of the students/PhD

---

[6] NOKUT is an independent expert body under the Ministry of Education and Research. The agency has a variety of expertise and is, for example, responsible for accreditation and external quality assurance in higher education.

[7] The Quality Assurance system was translated into English in 2010. I am citing this version because this was the applicable English version on the university's webpage during the time period of the PhD project. An observant reader might notice that I refer to the 2012 version of QAS in the papers, which was the current version during the data collection. However, the quality description was the same both in both the 2010 and 2012 versions.

candidates or, in other words, all conditions of significance to their learning, for health and welfare.

Quality indicators for teaching and learning

Student evaluation is one of many quality indicators for teaching and learning. Examples of other indicators or measures of quality teaching and learning are academic achievements like assessment results and student perseverance—moreover, evaluations from an academic's perspective like self-reported teacher evaluations and peer observations. Educational evaluations can comprise multiple indicators for educational quality (Ackerman et al., 2009; Alderman et al., 2012; Berk et al., 2005; Harvey & Newton, 2004; Trigwell et al., 2012).

Balance between Quality Assurance and Quality Enhancement

Among the most comprehensive and frequently cited definitions of quality assurance and quality enhancement are the ones listed in the Analytic Quality Glossary (Harvey, 2004-20). Assurance of quality in higher education is a process of establishing stakeholder confidence that provision (input, process and outcomes) fulfils expectations or measures up to threshold minimum requirements. Quality enhancement is a process of augmentation or improvement.

Quality assurance and quality enhancement are distinct but related activities and complex phenomena (Williams, 2016) that vary in definition and understanding according to time, context and to different actors or stakeholders (Harvey, 2007). John Biggs (2003) referred to quality enhancement as *prospective,* and as the improvement of quality by continually striving to improve teaching and learning, and quality assurance as *retrospective* in assuring quality by requiring conformity to externally imposed standards.

Student evaluation has been torn between the conflicting discourses of consumerist-driven quality assurance (what students want) and academic quality enhancement (what the students need to effectively learn) (Bowden & Marton, 1998; Darwin, 2016).

Quality work

In recent years the term "quality work" has been introduced as a concept that aims to integrate different internal processes involved in quality enhancement—formal and informal, organisational and pedagogical dimensions of quality—when studying what contributes to quality enhancement (Elken & Stensaker, 2018; Elken & Stensaker, 2020b). The definition of

11

quality work is "activities and practices within higher education institutions that address the quality of its educational provision" (Elken & Stensaker, 2018, p. 190), as an umbrella concept that focuses on informal and routine work as well as formal organisational structures created by different actors (Elken & Stensaker, 2020b). In a recent systematic literature review on quality initiatives in higher education, Bloch et al. (2020) conclude that there are many types of quality work practices and many conditions that can influence quality and that there is a need for more knowledge about the effects of quality work. In quality work, both pedagogical and organisational dimension are regarded as important "and should be seen as distinct but related dimensions of quality in higher education" (Elken & Stensaker, 2020b, p. 14). Student evaluation is one of many practices of quality work within institutions that address quality. In this thesis both pedagogical and organisational dimensions of student evaluation are explored.

2.0 Literature review

There are different approaches to and purposes of literature reviews (Boote & Beile, 2005; Grant & Booth, 2009; Randolph, 2009). Maxwell (2006) emphasises that the most important aim of literature reviews for doctoral students is to identify the relevance of research literature to one's own study. This is supported by Randolph (2009), who also suggests that a dissertation review has multiple goals. In this section, I aim to provide an overview of the field of research wherein my project is situated, not an exhaustive review "locating every available piece of research on a certain topic" (Randolph, 2009, p. 3), but to identify *relevant* literature related to the overall research question: in short, what characterises research on evaluation and particularly what we know and do not know about student evaluation practice and its uses from different actors' perspectives. Moreover, I aim to critically analyse previous research and identify central issues in the field.

Within research literature and databases student evaluation is labelled with many different key words and terms like "student feedback", "student rating", "student course evaluation", "student evaluation of teacher performance" and "student evaluation of teaching". I would like to recognise that I, in the three papers and this extended abstract, use the terms "student evaluation of teaching", "student evaluation" and "student course evaluation" interchangeably. I conducted several literature reviews from 2016–2020 and used the terms above in combination with each other (combined searches) and with relevant other terms to delimit and specify my

searches. The literature reviews are primarily conducted in the database Education Resources Information Center (ERIC), the most comprehensive database for pedagogics and educational research. I also included papers from the reference list of particularly relevant articles retrieved from searches in ERIC. I included peer-reviewed research from higher education only, excluded papers from primary, secondary and high school education. Additionally, I included some Norwegian research reports and white papers relevant to this project, as well as book chapters from (scholars in) the field of academic development. The literature reviews focused on research outcomes or findings, which is probably the most common focus (Randolph, 2009). However, I also analysed the dominant research methods and the origin of research. The literature reviews in the initial phases of the project were conducted as background for establishing the aims and research questions of the study. During the writing of the papers, new literature reviews relevant to the aims of the papers were conducted. I am in this section of the thesis presenting an updated literature review (conducted until December 2020) guided by the overall research question to position my research.

Different countries and institutions have different ways of organising evaluation and quality assurance. The most frequent format of student evaluation is surveys sent to the students at the end of the courses or programmes (Alderman et al., 2012; Richardson, 2005). The content of these varies, and they have been described as student satisfaction and student experience surveys (Klemenčič & Chirikov, 2015). In Anglo-Saxon countries, it is common to use standardised surveys for evaluation of teaching/teacher performance (SET) and course and programme evaluations; in North America are these often separate surveys. At the university where this study is carried out, course and programme evaluations also comprise questions that evaluate the teachers, meaning that there are not separate student evaluation surveys of each teachers' teaching. As educational contexts differ between countries, and most published research on student evaluation in English accessible in educational databases is from Anglo-Saxon countries, these findings are not necessarily transferable to other countries. Moreover, research on student evaluation from European countries closer to the Norwegian context is often published in the original language in the country of the study, and therefore not as accessible as English publications. The literature reviews in this thesis are conducted in English. However, I have included some texts in Scandinavian languages: Norwegian, Swedish and Danish. These are mainly derived from reading papers in English written by Scandinavian scholars who referred to research published in Scandinavian languages in reference lists.

I would like to note that the evaluation terminology and descriptors dominating in the databases seems to refer to evaluation as instruments that are inherently rationalistic. When research on evaluation of teaching is designed to assess "customer satisfaction", "teaching effectiveness", "teacher performance", "teacher instruction" and more seldom as "processes facilitating student learning", "interaction between teacher and students" the terminology indicates that teaching can be understood as a product that teachers deliver to students as customers. I understand the terminology in itself as value-laden and not neutral with reference to a rationalistic world view in which teacher ratings are described as measures of teaching effectiveness[8] (e.g., Uttl et al., 2017). Returning the reader's attention to the Scandinavian context, Scandinavian scholars in higher education have questioned the understanding of educational quality as similar to that in business and industry (Dahler-Larsen, 2019; Stensaker, 2007), particularly with respect to judging educational quality by effectiveness measures. Thus, Scandinavian scholars suggest that the relationship between quality and effectiveness is complex (Bleiklie & Frølich, 2014; Skodvin & Aamodt, 2001; Stensaker & Maassen, 2001). In the conducted literature reviews, I found several Scandinavian educational policy studies in the area of quality assurance (e.g., Bergh, 2015; Bleiklie & Frølich, 2014; Gornitzka et al., 2004; Stensaker, 2006; Stensaker et al., 2019) but very few empirical studies on student evaluation use from Scandinavian countries (e.g., Andersen & Søndergaard, 2006; Edström, 2008; Nørholm, 2008). Consequently, I consider this study a contributor to a better understanding of evaluation practice in Scandinavia.

2.1 Historical glimpse of student evaluation

Precursors to student evaluation have been found to have existed at Medieval European universities (Knapper, 2001), where students' committees were appointed by the rector to "assure teachers adhere to defined orthodoxies" (Darwin, 2016, p. 4). A more modern appearance of student evaluation dates to the 1920s, to Purdue University in the United States and the student feedback questionnaire *Purdue Rating Scale for Instructors*. This student evaluation survey intended to assess student opinions of the teaching and the learning process, which in turn could be used in teaching improvement for the individual teacher (Marsh, 1987). Remmers, who introduced the rating scale, later expressed that academics should be "cautious about the use and interpretation of student ratings indicating that his Purdue scale was not

---

[8] The debates about what student evaluation really measure and the definitions of teaching effectiveness are complex and inexhaustible and will just briefly be discussed in this thesis, for a more comprehensive understanding, see e.g., (Abrami et al 2007, Bedgood and Donovan 2012).

designed to serve as a measure for teaching effectiveness" (Marsh, 1987, p. 258). Student evaluation of teaching did not spread rapidly until the 1960s when more formal evaluation systems were established at universities (Centra,1993), yet it was still considered to be for academic use and was voluntary (Darwin, 2016). The establishment of formal student evaluation systems took place in Europe about twenty years later than in America (ibid). In many European countries, such systems were established in the 1980s when student numbers and budgets increased and politicians wanted more control over how public money was spent as part of quality assurance for the purpose of accountability and control (Westerheijden, Hulpiau, et al., 2007). At the same in the 1980s, the public sector in many Western countries, including Norway, changed in a wave of many business-inspired reforms described as New Public Management (NPM) (Bleiklie, 1998; Bleiklie & Frølich, 2014). These reforms valued management techniques like measurement, auditing, evaluation, incentives and sanctions for the sake of increased efficiency (Christensen et al., 2020). Within NPM, evaluation can be understood as a management technique central to auditing processes and measurement of educational quality. An understanding of educational quality as a *product* and a belief that it can be measured seemed to have been strengthened during the decades where NPM reforms were introduced in the public sector (Dahler-Larsen, 2019; Gulbrandsen & Stensaker, 2003). Public sector reforms building on principles from the private sector have had an impact on how higher education in Norway today is steered, e.g., with stronger public steering and introduction of auditing processes and external evaluations than previously (Bleiklie & Frølich, 2014). Nonetheless, quality, and particularly educational quality, is recognised as more than a product and as a complex *process*—a recognition that is incorporated in the current description of educational quality from NOKUT (2016). How quality is conceptualised in external quality assurance is closely related to the development of internal quality culture (and an internal assurance system) (Danø & Stensaker, 2007). More recently within the concept of "quality work", informal internal quality initiatives articulated are also important for maintenance and enhancement of educational quality (Elken & Stensaker, 2018; Elken & Stensaker, 2020a).

Quality assurance and educational evaluation in European higher education is today based upon The Standards and Guidelines for Quality Assurance in the European Higher Education Area (ESG), created by the European Network for Quality Assurance in Higher Education (ENQA), a network that was established in 2000. Most European countries have established evaluation practices that comply with ESG, and included student evaluation as part of quality assurance systems (Stensaker et al., 2007). This is also the case for Norwegian higher education

(Studiekvalitetsforskriften, 2018). External regulations can be seen as a strong driver for implementation of new practices and one reason why student evaluation as part of quality assurance has become ubiquitous in higher education.

This short historical glimpse shows that use of student evaluation has changed since it was introduced in higher education, from internally self-driven voluntary tools to improve teaching to measures that are also externally regulated for control and accountability (Bamber, 2011).

2.2 Quantitative studies dominate the research of student evaluation

Prior research is dominated by quantitative studies that investigate bias, validity and reliability of evaluation methods/surveys, response rates, response styles etc. I will in this section present some of the findings about bias, validity and reliability as this knowledge is relevant in understanding what characterises student evaluation, but also because statements about bias dominate the debate about student evaluation. This debate seems to be full of myths, claims and anecdotal references, and not based on research. Evaluation seems to be a phenomenon that "everybody" has an opinion about and "knows how to do" without having studied the field comprehensively[9].

Several studies conclude that student evaluation surveys are valid and reliable instruments to measure teaching effectiveness (Ginns & Barrie, 2009; Lemos et al., 2011; Marsh & Roche, 2000; Socha, 2013; Wright & Jenkins-Guarnieri, 2012; Zhao & Gallant, 2012) and suggest these are more valid and reliable measures of teaching quality than any other measure when properly designed (Berk et al., 2005). Other have questioned the validity and reliability of SET (Curby et al., 2020; Dommeyer et al., 2002; Spooren et al., 2013). Studies that question the validity of SET relate it to the low correlation between high ratings and student learning (Kember & Wong, 2000; Uttl et al., 2017). When reading studies about validity and reliability of evaluation, the reader should keep in mind the plethora of existing evaluation surveys in the sector, and that the cited research often consists of validity and reliability studies of single instruments. Hence, the findings apply to these instruments and not the broad spectrum of evaluation methods.

---

[9] this claim is supported by analysis of American university papers and essays (Feldman, 2007; Linse, 2017).

Many researchers have investigated how students respond to evaluations and whether their responses are biased. After the transfer from written in-class evaluations handed out by teachers to online surveys distributed by administrative staff (Estelamani, 2015) the response rates declined (Adams & Umbach, 2012; Dommeyer et al., 2004; Lipsey & Shepperd, 2020; Sax et al., 2003) and the responses became less constructive and more negative (Gakhal & Wilson, 2019). Low response rates with low representation from the total student group implies that the response rate is a bias in itself (Bacon et al., 2016; Curran, 2020; Reisenwitz, 2016). Adams and Umbach (2012, p. 585) suggested that evaluators should pay attention to nonresponse bias which they emphasise "occurs when patterns of nonresponse exist across particular groups". While it is possible to adjust evaluation practice in order to facilitate increased response rates, e.g., putting evaluation on the agenda or timetable (Young et al., 2019), sending students reminders (Bennett & Nair, 2010; Dommeyer et al., 2004), or giving feedback to the students about actual use (Nulty, 2008), there exist several biases that the teachers can not affect. These are, among others, biases that concern gender, ethnicity, class size, weather conditions, elective courses and discipline.

Female teachers receive poorer evaluations than male teachers (Boring, 2017; Fan et al., 2019; Mengel et al., 2019; Sprague & Massoni, 2005). White teachers get better ratings than teachers of African or Asian decent (Basow et al., 2013; Chávez & Mitchell, 2020). Some subjects tend to get poorer ratings than others, e.g., courses in statistics are rated lower than courses in the humanities (Davies et al., 2007; Uttl & Smibert, 2017). Students in smaller classes evaluate courses better than students in bigger student groups (Braga et al., 2014; Davies et al., 2007; Liaw & Goh, 2003; McPherson, 2006), elective courses are evaluated more favourably than compulsory (Feldman, 2007; Patrick, 2011). Teachers' attractiveness and personality also affect students rating, indicating that "conventionally attractive" (Hamermesh & Parker, 2005; Wolbring & Riordan, 2016) and extroverted, open, agreeable and conscientious teachers (Patrick, 2011) get better evaluations than those who do not possess these characteristics. Students tend to give lower ratings on rainy days than sunny days (Braga et al., 2014). This could be related to what Grimes et al. (2017) described as affective evaluation; the fact that how students feel during and about the course rather than what they think (i.e., cognitive judgement) may affect their evaluation responses.

There are also biases that teachers (consciously and unconsciously) can control or affect; teachers who "give easy grades" are rated higher than those who are not that lenient in their grading (Carrell & West, 2010; Langbein, 2008; Patrick, 2011; Weinberg et al., 2009); the same is the case with those who provide "service" or treats to students (Hessler et al., 2018) and for teachers who present the course material in a well-organised manner (Donnon et al., 2010). Courses that require more effort and workload from students receive poorer evaluation than those requiring less (Braga et al., 2014; Donnon et al., 2010).

2.3 Different purposes and users of student evaluation in contemporary universities

In the literature, student evaluation is described as having many potential users and different purposes. The purposes of student evaluation are often referred to as a dichotomy, like audit and development (Edström, 2008), accountability and improvement (Bowden & Marton, 1998), appraisal and developmental purpose (Kember et al., 2002) and, the most frequently used dichotomy, quality assurance and quality enhancement (e.g., Biggs, 2003). Additionally, for these purposes, student evaluation is regarded as an aspect of student empowerment and as helpful in selecting future courses/programmes for students (e.g. Bennett & Nair, 2010). For teachers, feedback on their teaching is important in order to improve courses and student learning (e.g., Ryan & Cousins, 2009); for academic developers, evaluations are drivers of enhancing educational quality and used when they "mediate between the institutional requirements for QA and academic norms" (Bamber & Anderson, 2012, p. 5); for administrative staff and at the institutional level, they are used for administrative purposes like tenure and appraisal, hiring/firing of academics, university rankings and accountability (e.g., Beran et al., 2005). Student evaluation is also essential in national accreditation processes, and in policy documents the overall functions are quality assurance and quality enhancement (e.g., Danø & Stensaker, 2007). In addition to the abovementioned purposes, student evaluation is mentioned as an outcome for use in research on teaching (e.g., Marsh, 2007). With this knowledge serving as background, I present in the next subsections research that says something about what characterises student evaluation practice today, what evaluation is actually used for and by whom.

2.3.1 Academics' perspective on student evaluation and use

Teaching academics are key stakeholders in student evaluation practices and are regarded as central users of student evaluation results. A general finding from research on the teachers' perspectives on student evaluation, including my first paper, is that teachers have overall positive attitudes towards student evaluation and value feedback from students (Beran & Rokosh, 2009; Borch et al., 2020; Nasser & Fresko, 2002; Stein et al., 2013; Wong & Moni, 2014). Student evaluation is central to quality assurance and studies on academics' perspectives on quality assurance have reported that academics are more sceptical of these processes and view them as "rituals", "games" or systems for control and accountability more than for quality enhancement (Anderson, 2006, 2008; Newton, 2000, 2002). Despite an overall positive attitude towards student evaluation, the direct use for improving their own teaching is found to be low (Beran & Violato, 2005; Beran et al., 2005; Kember et al., 2002; Nasser & Fresko, 2002; Stein et al., 2013), yet others report that teachers use student feedback in teaching preparations (Moore & Kuol, 2005; Safavi et al., 2013) and that male teachers adjust their teaching more than females based on students' responses (Kogan et al., 2010). Studies have also found that teachers are sceptical of student evaluation (Balam & Shannon, 2010), particularly in overemphasising student evaluation results as the only source of reference when judging teaching effectiveness and quality (Ackerman et al., 2009; Anderson, 2006; Beran & Rokosh, 2009; Berk et al., 2005; Vasey & Carroll, 2016).

There are some known aspects of why teachers use student evaluation for teaching improvement to only a minor extent. One relates to the abovementioned multiple purposes of evaluation and a belief that student evaluation is mainly conducted because of accountability and control (Harvey, 2002). Other reasons for low use relate to the students' response styles and how they do not respond to surveys about teaching in "mindful" ways (Dunegan & Hrivnak, 2003) they have little knowledge about pedagogics and regard what constitutes "good", "effective" or "quality"[10] teaching differently from teachers (Ackerman et al., 2009; Anderson, 2006; Nasser-Abu Alhija, 2017). Further, researchers have found that little support and help with interpretation of findings is correlated with low use of findings in teaching and course improvement (Penny & Coe, 2004) as many teachers have limited knowledge about data analysis (Boysen et al., 2014; Datnow & Hubbard, 2016) or have negative feelings about

---

[10] See e.g., (Nasser-Abu Alhija, 2017) about how these terms are used interchangeably in literature about student learning without having agreed upon a definition for the term.

negative student evaluation responses that make then opposed to the results (Kogan et al., 2010; Nasser & Fresko, 2002). Other aspects of student evaluation that may hinder teachers' use of the data in course improvement are knowledge about low content validity in measuring teaching effectiveness (Dunegan & Hrivnak, 2003; Zabaleta, 2007), surveys conducting judgmental information about teaching (Edström, 2008; Kember et al., 2002) and about student satisfaction (Arthur, 2020; Bedggood & Donovan, 2012) or superficial surveys with unclear rating criteria (Edström, 2008; Schuck et al., 2008) rather than asking for feedback about aspects that can be used for improvement (Hauer & Papadakis, 2010; Kogan & Shea, 2007; Snell et al., 2000). Low correlation between student learning and teacher ratings (Clayson, 2009; Galbraith et al., 2012; Kember & Wong, 2000; Uttl et al., 2017) is also a reason why some scholars recommend using student evaluation with great care or even consider abandoning these (Uttl et al., 2017). Despite some hindrances to the use of student evaluation results for formative purposes, academics have experiences with meaningful student evaluation practice that benefit their professional development. Academics report that they find student evaluation useful (Hammer et al., 2018; Stokke et al., 2019), e.g., in reflective practice and professional development (Moore & Kuol, 2005; Niessen et al., 2009). Moreover, the student ratings did increase for teachers who participated in academic development activities (Trigwell et al., 2012) and for those who actively used student feedback to reflect upon their teaching (Golding & Adam, 2016; Winchester & Winchester, 2014). Formative student evaluation (Andersson et al., 2012; Siu, 2012; Youssef, 2017) and mid-term evaluations (Veeck et al., 2016) are found to increase teachers' reflections on their own teaching and are a viable complement to end-of-semester surveys. Evaluation during courses invite students to provide information about aspects of courses that teachers can use for improvement of teaching before the course is over and is therefore recommended if the aim is to improve teaching (Beran & Rokosh, 2009; Huxham et al., 2008; Ramsden, 2003; Vasey & Carroll, 2016).

Scholars within the field of education and evaluation suggest using separate evaluation approaches for the purposes of quality improvement and quality assurance (Barrie & Ginns, 2007; Patton, 2008). Nevertheless, academics expressed that one evaluation method is often intended to be used for different purposes which can lead to confusion, uncertainty and low use among the academics (Edström, 2008; Hulpiau et al., 2007; Huxham et al., 2008; Leth Andersen et al., 2013).

2.3.2 Organisational dimensions and administrative use of evaluation

Administrative staff and university managers are key stakeholders in student evaluation practice and central to the development of quality assurance systems (Michelsen & Aamodt, 2007; Newton, 2000). Administrative staff or administrators of evaluation, including deans and managers, have more trust in student evaluation as a valid measure of teaching effectiveness than academics (Morgan et al., 2003; Shao et al., 2007). The academics have little influence on evaluation procedures (Dommeyer et al., 2002), yet those who develop and manage quality assurance systems decide, for example the timing of when evaluation is carried out. With the change in format from paper to online surveys several potential benefits are suggested: the time frame for response is more flexible and convenient (Donovan et al., 2010), it is easier and more time-efficient to administer (Mau & Opengart, 2012) and enables students to comment anonymously and more comprehensively (Burton et al., 2012; Gakhal & Wilson, 2019). While paper evaluations in classroom settings often were handed out by the teachers and responded to immediately, today's online surveys are sent to the students by administrators, who also set the time frame for response (Estelamani, 2015). Despite low response rates and other disadvantages, student evaluation results are used in contemporary universities in personnel decisions like promotion, pay increments and tenure (Ahmadi et al., 2001; Kember et al., 2002). A Canadian study by Vargas-Madriz et al. (2019) explored how student evaluation was used by the administrators and found that administrators were aware of the issues with bias and validity concerning these evaluations and strived/attempted to get contextualised understanding about the courses in addition to the SET scores. Contradictory findings were identified by other scholars who stated that administrators of evaluations and academics were unaware of the vast research on student evaluation (Linse, 2017), and had little competence about how to interpret statistical data (Boysen et al., 2014; Mitry & Smith, 2014). A research project from New Zealand has focused on organisational dimensions of student evaluation practice and found that staff engagement in student evaluation processes were influenced by different aspects with the operationalisation of the evaluation systems (Stein et al., 2012; Stein et al., 2013). Among other findings, the study found that by making evaluation infrastructure easier and faster (Moskal et al., 2016) and being clear about the institution's expectation of using evaluation for professional development (Stein et al., 2013), academics' engagement with student evaluation increased. A systematic implementation of quality assurance systems that encourages staff to use evaluation for educational improvement has been documented to affect evaluation results positively (Barrie et al., 2005).

2.3.3 Student evaluation in documents

It is well known that quality assurance processes in general involve time-consuming paperwork and documentation. This can be considered as an organisational dimension of quality work. In research on student evaluation and quality assurance in higher education evaluation, reporting is often mentioned without exploring what these processes actually contribute. Regulation of higher education in Norway states that each institution shall have established quality assurance systems and *document* how they assure and enhance educational quality (Studietilsynsforskriften, 2017). In reviewing the literature, little was found directly related to documentation routines in student evaluation processes/of student evaluation. However, I found one report from 2009 (Froestad & Haakstad, 2009) that analysed institutional educational quality reports from 17 institutions in 2003-2007 conducted on behalf of NOKUT. Further, documentation of educational quality was a sub-topic in two studies exploring academic leadership and academics' perspectives on managerialism in Norwegian higher education (Ese, 2019; Johansen, 2020). Both studies showed that academics considered the evaluations, particularly the reporting of findings, as time-consuming administrative and bureaucratic processes that did not promote educational quality but were conducted because of requirements. Johansen (2020, p. 166) described that academics considered these activities to be "proforma". I did not find any studies that analysed how student evaluation practice was described in educational quality reports or how the documents themselves contributed to the understanding of evaluation as non-human actors of evaluation processes. As documentation is central to evaluation practice, I aimed to explore this further in order to answer the overall research question in the thesis. While several papers point at the high number of documents produced in evaluation processes, I found only a small number of studies that contemplate the contribution of such (Dahler-Larsen & Krogstrup, 2001; Ryan, 2015). Nevertheless, in methodology literature, documentary studies and organisational communication studies grounded in constructivism are documents described by several scholars as artifacts in organisational activities (Atkinson & Coffey, 2011; Castello & Inesta, 2012) and/or as non-human actors affecting organisational processes and realities (e.g., Cooren, 2004; Dahler-Larsen, 2011; Prior, 2003, 2008; Putnam & Nicotera, 2009; Scott, 1990). Because of the sparse literature on the topic from higher education, I rely on these perspectives as analytical approach in the third paper.

2.3.4 Characteristics of student evaluation practice from students' perspective

Students are respondents or the evaluand in evaluation processes; therefore their perspective on student evaluation is of high importance for the development of evaluation practice. However, in the early years of research on student evaluation, little attention was given to their perspective (Darwin, 2016). In recent years, more researchers have included the student perspective and we have gained more knowledge about different aspects of evaluation like students' motivation to provide feedback (Hoel & Dahl, 2018) and their experiences with student evaluation practice (Heine & Maddox, 2009; Spencer & Schmelkin, 2002). When asked about their attitude towards evaluation, studies found that students are generally willing to provide their feedback about courses (Ahmadi et al., 2001; Brown, 2008; Campbell & Bozeman, 2008; Kite et al., 2015; Spencer & Schmelkin, 2002), and they consider evaluations important in student empowerment (Hanken, 2011). Students state that they take evaluation seriously and provide honest and fair answers[11] (Ahmadi et al., 2001; Brown, 2008; Kite et al., 2015; McClain et al., 2018). The students seem to be motivated to participate and provide feedback as long as they believe their responses matter (Ernst, 2014; Hoel & Dahl, 2018), and some groups are more likely to respond than others. Female students are more likely to respond than male students (Avery et al., 2006; Porter & Whitcomb, 2005; Sax et al., 2003). Those who expect high academic achievement measured by grade are responding more often than those who expect lower grades (Avery et al., 2006; Porter & Umbach, 2006; Porter & Whitcomb, 2005; Sax et al., 2008). In the debate about student evaluation many academics often explain low response rates with 'survey fatigue' (Adams & Umbach, 2012; Bennett & Nair, 2010) but counter evidence from other scholars has suggested that students are not over-surveyed (Stein et al., 2020; Wiley, 2019). Low response rate has rather more to do with lack of feedback about what teachers do with student evaluation results (Harvey, 2003; Leckey & Neill, 2001), and a belief that their opinions do not matter. Students are unsure about whether their feedback is taken seriously and about the actual use (Brown, 2008; Leth Andersen et al., 2013; Spencer & Schmelkin, 2002) and some believe their opinions do not matter (El Hassan, 2009). This may be related to the lack of feedback about their responses and how the university plans to use the evaluations (Leth Andersen et al., 2013; Stein et al., 2020; Wiley, 2019). Students' motivation is also related to whether they are given the opportunity to provide meaningful feedback that they believe will contribute to improvement of teaching, course content and format (Chen &

---

[11] Research on bias in students' evaluation responses may complement/refine these statements and is presented in subsection 2.1.

Hoshower, 2003), and are less positive towards evaluation if they believe their responses are used for personnel decisions (McClain et al., 2018). Some studies have found that students consider standardised surveys to be inadequately contextualised (Leth Andersen et al., 2013; Wiley, 2019), that they suggest including questions that evaluate the relationship between courses (Kogan & Shea, 2007; Varhaug, 2012), increase dialogue about the results (Spencer & Schmelkin, 2002) and emphasise the need for mid-term evaluations (Leth Andersen et al., 2013; Veeck et al., 2016; Wiley, 2019; Winchester & Winchester, 2012). Students are more positive towards mid-term evaluations than evaluations at the end of courses because the former benefit them while they are enrolled in courses (Leth Andersen et al., 2013). Some of the abovementioned limitations of student evaluation practice are also expressed by academics, like superficial surveys (Edström, 2008; Leth Andersen et al., 2013), fragmented course evaluation surveys (Kogan & Shea, 2007) and little dialogue about evaluation results (Leth Andersen et al., 2013).

Studies that analysed how students and academics valued the importance of different aspects of evaluation found that students and academics regarded the importance of the dimensions differently (Braga et al., 2014; Feldman, 2007; Fischer & Hänze, 2019; Hadad et al., 2020). These findings underlie the suggestion expressed by academics that student ratings should not be the only measure of teaching quality (Ackerman et al., 2009; Ballantyne et al., 2000; Benton & Young, 2018; Berk, 2018; Marsh, 1987, 2007).

2.3.5 Scandinavian studies on student evaluation practice and evaluation use

Much of the recently published research is challenging and points to limitations with today's evaluation practice; this is also the case for Scandinavian studies conducted in educational contexts that are more similar to the context of this study than studies from Anglo-Saxon countries (e.g. Bergfjord, 2014; Edström, 2008; Gynnild, 2002; Hanken, 2011; Leth Andersen et al., 2013; Roxå & Bergström, 2013; Von Müllen, 2006; Wiers-Jenssen et al., 2002). A commonality between several of the Scandinavian studies is that they point at limited use of student evaluations for teaching and course improvement and rather for symbolic use (Ese, 2019; Nørholm, 2008), as a ritual (Hanken, 2011; Moldt, 2006), a "fire alarm" function (Edström, 2008) and quality assurance (Andersson et al., 2012; Johansen, 2020). Moreover, academics request more feedback from and discussions with their leaders about the intentions with evaluation practice (Leth Andersen et al., 2013; Moldt, 2006), and collegiality in

interpreting the results (Gynnild, 2002; Roxå & Bergström, 2013; Varhaug, 2012). Research referred to above on student evaluation is characterised by qualitative studies exploring evaluation in the local university context and show a variety of locally developed methods. Although the majority of Scandinavian research on student evaluation seems to be empirical, qualitative studies, Scandinavian scholars have also published papers literature reviews (Aarstad, 2012; Strømsø, 2016) and quantitative studies, e.g., of students' response styles (Bergfjord, 2014). More recently, Scandinavian researchers have been pointing to a need to explore initiatives promoting educational quality, like student evaluation, in an even bigger and complex context where different perspectives and dimensions are integrated in the field of quality work. In quality work, more knowledge about the interaction between different actors, between policy and practice and pedagogical and organisational dimensions is requested (Elken et al., 2020; Elken & Stensaker, 2018).

2.4 Pedagogical dimensions of student evaluation practice

As the literature review above shows, students and teachers suggest improved teaching and student learning as a central purpose of student evaluation. Nonetheless, a recent meta-analysis and re-analysis of prior published meta-analysis by Uttl et al. (2017) found little or no correlation between highly-rated professors and student learning. Moreover, they found weaknesses with both methods and analysis of the earlier meta-analysis. Based upon the weak relationship between student learning and how they rate teachers, scholars have questioned what we really are measuring with student evaluation and the validity of traditional surveys (Bedggood & Donovan, 2012). Some have suggested that students be understood as consumers (Little & Williams, 2010; McCulloch, 2009) or customers (El Ansari & Moseley, 2011; Mark, 2013; Moldt, 2006; Saunders, 2014; Sharrock, 2000), reflecting the increased market orientation of higher education (Anderson, 2006), for instance when they are asked to rate their satisfaction with teachers and teaching. Moreover, researchers have stated that many of the surveys are teacher- and teaching-focused rather than student- and learning-oriented (Andersen & Søndergaard, 2006; Blackmore, 2009; Bovill, 2011; Edström, 2008; Von Müllen, 2006). Consequently, in the last two decades scholars have provided new approaches, models and methods to conduct student evaluation that facilitate feedback from students about their learning and suggested that such evaluation results are better suited for teaching improvement

and course development than the traditional and often teaching-oriented surveys[12], yet the most-used student evaluation format is still surveys distributed in end of courses (Richardson, 2005). Examples of more learning-related and student-centred written evaluation approaches are Competence-based evaluation developed by Bergsmann et al. (2018; 2015) with support from the European Agency for Quality Assurance (ENQA), Contextualised evaluation by Nygaard et al. (2011), Fourth generation inspired evaluation approach by Darwin (2012) and improvement driven evaluations by Golding and Adam (2016) and Roxå et al. (2021). Many scholars have also suggested principles that should be considered when developing a more student-centred evaluation practice based upon empirical research showing that evaluations are rather teacher-oriented (Andersen & Søndergaard, 2006; Christensen, 2010).

Examples of some dialogue-based student learning-centred evaluation approaches: peer-assisted evaluation by Carbone et al. (2015), a feedforward, concurrent and feedback approach by Catchard et al. (2014), collaborative midterm evaluations by Veeck et al (2016), students as peer observers (Huxham et al., 2017), the Student Enhanced Learning Trough Effective Feedback model (Freeman & Dobbins, 2013) and different participatory evaluation approaches (DeLuca et al., 2009). Focus groups have also been suggested as supplements to traditional written evaluation methods (Berk et al., 2005; Brandl et al., 2017; Varga-Atkins et al., 2017). Other scholars have contributed to the discussion about which evaluation methods best facilitate dialogues about student learning processes, have described principles of qualitative evaluation methods and have suggested using these as supplement to traditional surveys without developing specific models (e.g. Andersson et al., 2012; Darwin, 2017; Huxham et al., 2008; Steyn et al., 2019; Tran, 2015; Tran & Nguyen, 2015). As students do not have the opportunity to be anonymous in dialogue-based evaluation approaches, some contemplate that student feedback from dialogue-based evaluations are not as candid as survey results (Afonso et al., 2005).

---

[12] Many of the existing surveys are described as learning-oriented, however, many of the newer approaches are developed because of a critique of traditional surveys. One example of a widely used standardised survey that aims to capture aspects of students' learning like deep or surface learning is the Course Experience Questionnaire (CEQ) developed by Paul Ramsden (2003).

2.5 System alignment: Incorporating student evaluation in constructive alignment

In an empirical study about student evaluation from Sweden, Edström (2008) found that educational policies regulating evaluation and actual evaluation practice were in conflict with each other. The intention with evaluation in the policy documents was that student evaluation should contribute to course development, but the actual student evaluation practice was rather teaching-focused and did not collect information suitable for teaching improvement and course development. Edström (2008, p. 105) argues that evaluation is a component of constructive alignment, or at least should be. She introduced the concept of "system alignment" as "a parallel to constructive alignment" by extending constructive alignment to the institutional level and including evaluation. By considering the university as a system, she described the system components in the concept as any macro-level structures like infrastructures, work process and policies. Within constructive alignment are learning outcome descriptions, teaching/learning activities and assessment aligned with each other and regarded as a premise for student learning (Biggs, 1999, 2003)[13]. Edström (2008, p. 105) described "constructively aligned course evaluation" as evaluation practices that "support the improvement of student learning".

2.5.1 Student evaluation in academic development

Student evaluation has been described by academic developers (AD)[14] as more than technical rational activities and rather as reflective practices (Bamber & Anderson, 2012; Bamber & Stefani, 2016). They also suggest embedding student evaluation more systematically in pedagogical practice in higher education institutions to improve teaching and learning (Edström, 2008; Ramsden, 2003; Roxå et al., 2021). Academic developers have suggested more active use of evaluation for formative purposes (Haji et al., 2013) and have pointed at the need to analyse what hinders evaluation being used in educational development (Saunders, 2011). Below I present possible explanations as to why these evaluations are not used more actively

---

[13] I have chosen to refer to the term "constructive alignment" (CA) as it is a well-known concept in the sector and also described in the local quality assurance system as something that should be evaluated (UiT, 2012). I would like to note that the concept of CA has been debated and criticised by scholars, e.g., by Ashwin et al. (2015), and Andersen (2010, p. 30) who suggested that CA as guiding principles in curriculum design can "simplify university pedagogics" and generalise courses in ways that inhibit deep learning and creative thinking.

[14] "Educational developer" and "academic developer" are European terms for academic staff employed in higher education institutions, typically in centres for teaching and learning aiming to support teachers in their pedagogical qualification and development. In North America the term for the same position is often "faculty developer". I am using the terms interchangeably in the thesis. For a review of global trends in academic development practices and an overview of academic developers' work and impact, including evaluation see e.g., (Gibbs, 2013; Sugrue et al., 2018).

in teaching improvement from ADs' perspectives, in addition to those presented in the literature review so far.

First, teaching-oriented surveys (Von Müllen, 2006) and summative course evaluations (Andersson et al., 2012) are found not to be suitable or valuable in teaching and course improvement. This has also been expressed by other educational researchers and resulted in suggestions of more learning-oriented evaluation approaches presented in section 2.4 above. Formative evaluation is also recommended in addition to summative surveys by ADs in order to obtain quality enhancement and, additionally, to quality assurance (Andersson et al., 2012).

Second, poor or non-existent feedback loops about evaluation results to teachers and students are suggested as a hindrance to using student evaluation in quality enhancement processes. Ensuring feedback loops in evaluation is suggested as an aspect of a more systematic approach to evaluation practice at universities, including plans about how to give feedback to the students on how the university is planning to act upon the student responses (Harvey, 2003; Leckey & Neill, 2001; Shah et al., 2017; Watson, 2003). The teachers need to be reminded of how the summative evaluations play an important role in quality assurance on the institutional level but also to see outcomes of the system "to fully understand the value of their input" (Andersson et al., 2012, p. 100).

A third aspect that is emphasised as a key element is support for interpretation of the evaluation findings (Wong & Moni, 2014). Different literature reviews and studies found that use of evaluation results in quality enhancement purposes increases if educational development units or ADs are involved in the development and interpretation of evaluation (Hampton & Reiser, 2004; Penny & Coe, 2004; Piccinin & Moore, 2002). Research has also indicated that there is a need for more knowledge about how to analyse evaluation data statistically. Scholars have published papers with principles about how to interpret quantitative evaluation data results (Abrami, 2001; Boysen et al., 2014; Franklin, 2001; Linse, 2017; Theall et al., 2001).

Studies in the field of academic development describe that ADs should take on the role of quality agents (Havnes & Stensaker, 2006) or change agents facilitating educational quality (Cordiner, 2014; Debowski, 2014; McGrath, 2020). Many of the studies have in common that several of the activities ADs are involved in are aiming to promote educational quality, such as support for implementing evaluation findings, creating collegiality about educational quality

issues (Ramsden, 2003), introducing formative evaluation approaches and involving students and teachers actively as partners in academic development. Incorporating student evaluations in a dyadic manner between teachers and students by regarding students as partners, peers, colleagues is a welcome development suggested by ADs (Cook-Sather et al., 2019; Felten et al., 2019; Mercer-Mapstone et al., 2017; Ramsden, 2003; Saunders & Williams, 2005). Key factors in establishing evaluation practice that are used to improve teaching, and hence student learning, are that teachers regard them as meaningful and have a sense of ownership over them, and that they are carried out with a scholarly and collegial approach (Ramsden, 2003).

Academic developers have pointed at the need for course and programme evaluations that go beyond asking evaluation questions that request feedback from students about whether or not teaching "worked" and rather about what, why and how it worked, and emphasised that evaluation must be seen as a continuous process and not a snapshot activity conducted after the programme delivery (Haji et al., 2013).

Many of the elements of this subsection that ADs have pointed to as important for success in establishing good evaluation practice helpful in improving teaching and learning are also found in literature about Evaluation Capacity Building (ECB)[15], sometimes under different labels or terminology, e.g., with actors referred to as "stakeholders", knowledge about how to analyse evaluation as "evaluation literacy" and partnership as "stakeholder involvement".

As this literature review has shown, student evaluation has many potential users and multiple purposes. Historically, the purpose of student evaluation has shifted from improvement of teaching to also incorporate quality assurance, accountability and student empowerment. Research on student evaluation practice is therefore situated in a tensional field between quality enhancement and quality assurance. Most research on student evaluation is from Anglo-Saxon countries with different higher education systems than in Scandinavia. The body of research has investigated fragmented parts of student evaluation, like aspects of student evaluation methods, response rates and bias with quantitative research methods or qualitative research from the perspective of a stakeholder group: students, teachers, leaders, etc. Much of this research can be considered fragmented, often de-coupled from formal organisational contexts, structures, educational policies and interaction between different stakeholders in student

---

[15] Evaluation Capacity Building is presented in paper two (Borch 2020) and in the discussion section of the thesis.

evaluation practices. Based on the literature review, I will argue that there is a need for more knowledge about evaluation in a Scandinavian context as well as research that explores how different actors interact in evaluation practice. As opposed to investigating fragmented bits and pieces regarding evaluation, this study's emphasis on evaluation *practice* incorporates analysis of different actors' perspectives on and participation in student evaluation practice. By exploring pedagogical and organisational aspects of evaluation practice, and in addition incorporating different stakeholders' perspectives and the interaction between them in the same study, I explore how student evaluation is carried out and embedded in the organisation. This study can be considered part of quality work[16] and aims to contribute new knowledge to the field.

3.0 Theoretical frameworks

In this chapter I present the theoretical (epistemological, ontological and analytical) frameworks that have guided the project. As mentioned in the introduction, this project draws on social constructivism[17]. Social constructivism is today regarded as a multi perspective of theories of science with roots in different philosophies of sciences which have in common that they questioned that knowledge is understood as purely objective and rational (Alvesson & Sköldberg, 2018). The different theories of sciences that social constructivism originates from are often described as constructivism. Lincoln and Guba (2013, p. 39) states that "Relativism is the basic ontological presupposition of constructivism", meaning that the world or entities in the world "exist only in the mind of the person contemplating them". This perspective does not reject the nature and the physical world but emphasises that knowledge is not "discovered" but "created" and that realities "depend on a transaction between the knower and the to-be-known in a particular context"; this is described as subjective epistemology (Lincoln & Guba, 2013, p. 40).

In addition to the theoretical framework of social constructivism, I have chosen three analytical perspectives in the interpretation and discussion of the empirical data in order to answer the

---

[16] The concept of quality work is presented in section 1.3 Educational quality.

[17] The terminology varies linguistically in literature between "Social constructivism" and "Social constructionism" (Alvesson & Sköldberg, 2018). I have chosen to use the term "Social constructivism", it is closer to the Norwegian term and the one that e.g., Guba & Lincoln use, however it should not be mixed with the understanding described by e.g., Piaget.

research questions. In qualitative research, theories are important for supporting the understanding and interpretation of empirical data beyond description (Malterud, 2015). Analysis within social constructivism is a practice of looking at the empirical material in a scientific way with support from theoretical concepts so that the social construction becomes visible (Esmark et al., 2005). The analytical perspectives I rely on are all rooted in constructivism: translation theory derived from Scandinavian new institutionalism, principles from fourth-generation evaluation theories, particularly Utilization-Focused Evaluation (UFE) and textual agency. These are presented below and in the papers.

3.1 Social constructivism

Berger and Luckmann (1966, p. 37), pioneers of social constructivism described the world of which we are a part as an "intersubjective world, a world we share with others". Social constructivism is based upon an assumption that social reality is constructed by social interactions and processes (Alvesson & Sköldberg, 2017; Järvinen & Bertilsson, 1998). Furthermore, that our understanding of the world and "the categories and concepts we use, are historically and culturally specific" (Burr, 1995, p. 3).

In our meetings with others, we "typify"[18] those we meet by categorising them based upon prior experiences. In addition to prior experiences are language, symbols and signs important when we give meaning to our surroundings (Alvesson & Sköldberg, 2018). When reality is regarded as socially constructed, "there is no single, observable reality. Rather, there are multiple realities, or interpretations of a single event" (Merriam & Tisdell, 2016, p. 9).

Berger and Luckmann (Berger & Luckmann, 1966, 2000) described how we continuously interpret reality through social interactions with others and that there is an inherently dependent relationship between the observer and what is being observed. For this research and my role as a researcher, this implies that I am part of the world I am studying. I will elaborate on my role as a researcher in the method sections.

Student evaluation in this project is understood as a social phenomenon and process constructed by the actors (non-human and human) in the organisation. Examples of non-human actors in

---

[18] "typify" is the word used by Berger and Luckmann (1966)

this study are documents, while human actors are students, academics, management and administrators. Student evaluation is *translated* and *institutionalised* by the actors involved in evaluation practice (Borch, 2020). The understanding of institutionalisation is rooted in social constructivism and in an understanding of habits and routines that are continually developed and formed through a certain way of acting, described by Berger and Luckmann as "habitualization". Through social interaction with others, we are also continuously establishing new understandings and ways of observing the world we are part of. In social interaction with others, "we habitualize and typify; these habitualizations and typifications-these habits, routines and categorizations-spread between actors, and as we do this, institutions, that is fixed patterns of thought and action, emerge: institutionalization occurs" (Alvesson & Sköldberg, 2018, p. 33). Alvesson and Schöldberg (2018, p. 34) state that "institutions are also represented by many other things, like linguistic symbols, physical artifacts, and so on. But only human enactment in roles makes the institutions, so to speak, come to life." An example of such institutionalisation is evaluation in modern organisations like universities. In this study, student evaluation practice is explored through the perspectives of students and academics, who make the institutions, like evaluation, come to life at programme level. I recognise that there are also other actors involved in evaluation practice, particularly from higher institutional levels and administrative staff, who also play a role when institutionalisation occurs. I have in addition included documents from other institutional levels, many of these are written by administrative staff. The documents—texts about evaluation practice—are written for an audience with an intention to be read by others. Documents may be regarded as linguistic symbols and physical artifacts that are given meaning through human enactment and interpretation. We can assume that the readers' interpretations are influenced by their prior experiences, values and perspectives, including me as a researcher. I acknowledge that both the analytical approach and my background influence how I give meaning to the documents; therefore, is it important to be being transparent about the analytical process and my background, this is elaborated on in Chapter 4.

Returning to the institutionalisation of phenomena or new ideas in organisations, like evaluation, I would like to underscore that any institutionalised phenomenon is originally created by humans as something external and objective before it is eventually perceived as something given (Alvesson & Sköldberg, 2018). This understanding of institutionalisation has been further developed and resulted in directions of institutional theory (Eriksson-Zetterquist, 2009). By institutionalised I refer to Krogstrup and Dahler-Larsen (2000, p. 285) who say that:

"when a standard is institutionalised, it means it is understood and referred to by many as the most appropriate, correct, effective, modern and taken-for granted way to organise" (my translation from Danish).

I base my understanding of documents on views expressed by scholars who emphasise documents as "social facts" that are dependent on human actors in order to be produced, read and acted upon (Atkinson & Coffey, 2011, p. 79). Documents are produced and used in "socially organized circumstances" (Prior, 2003, p. 16) and "do things" with organisations and shape organisational realities (Cooren, 2004). Prior (2003, p. 4) underscored that "documents are not just manufactured, they are consumed". Documents function and are used in different ways by different readers; Prior (2008) suggests therefore that researchers should study how the documents function in their social settings rather than simply what they contain. Aligned with Prior's suggestion, and what I regard as a constructivist understanding of documents, I am studying how documents function and are used by human actors and from a textual agency perspective. I have included documents and interviews with academics and students in this project in order to better understand how student evaluation is carried out and used at the university. However, I acknowledge that documents are not "transparent representation of organisational routines, decision-making processes or professional practices" (Atkinson & Coffey, 2011, p. 79). Therefore, documents alone would not have been enough to achieve the aim of the thesis. The interviews with students and academics and the documents are complementing each other in portraying what characterises evaluation practice at the university. Social constructivist analysis aims to understand how the social construction of practice and processes are established through social interaction between subjects and objects (Esmark et al., 2005). Language is central to social interaction and our experience of the world, and therefore of great interest in social constructivism (Burr, 1995, 2015). Burr (2015, p. 72) gives several examples of how words may have different meanings to people depending on context and the people who speak or write the words and the ones who interpret them, hence she states that "language constructs rather than represents the world". In the document analysis in this project, I analyse the content of the documents but also how the documents are given meaning, how they function and are given agency by actors at the university.

3.2 Evaluation use and Utilization Focused Evaluation (UFE)

Evaluation theories and how evaluation is understood have changed over time and are highly influenced by the time we live in (e.g., Christie & Alkin, 2008). It is neither my intention, nor indeed possible, to present all paradigms and generations in evaluation theories in this thesis. In short, the approaches to evaluation I am referring to are grounded in what Guba and Lincoln (1989) described as a fourth-generation evaluation paradigm, rooted in constructivism and developed as a response to previous positivist-oriented evaluation models. Since then, several approaches and theories have been developed based upon a constructivist understanding of evaluation as a social phenomenon. This includes a stronger emphasis on evaluation processes and stakeholder involvement and a broader understanding of evaluation use than earlier approaches to evaluation. One example is Utilization Focused Evaluation (UFE) which is the analytical framework of the first paper. Different approaches to understanding evaluation use, including UFE, are presented below.

As mentioned earlier in this extended abstract, Kirkhart (2000, p. 6) concludes in a meta-analysis of evaluation use that there exist many definitions, understandings and types of evaluation use; she argues that the term is "inadequate and imprecise" and therefore proposed the term "influence" *in extension to* "evaluation use". I share her understanding of evaluation influence and acknowledge that evaluation results can gradually emerge over time during and after an evaluation is conducted (Kirkhart, 2000). Evaluation influence incorporates the enlightenment function of evaluation and increased evaluative thinking. Some of the influences are described as unintended or unaware impacts of an evaluation. From my point of view the term "evaluation influence" is illustrative of how evaluation use in fourth-generation evaluation is regarded in much broader terms than earlier understandings.

Evaluation use has been a concern among evaluation professionals over many decades. It is probably one of the most-studied aspects of evaluation (Henry & Mark, 2003). When this was problematised in the literature at first, it was based upon an assumption that evaluation results and reports often were not read and used the way the evaluators intended (Preskill & Torres, 2000). The assumption was followed by research on use of evaluation and theories about evaluation use were developed (Alkin & King, 2016; Alkin & King, 2017; King & Alkin, 2019). Rich (1977) distinguished between three types of evaluation use: instrumental, conceptual and symbolic use. These three kinds of use are still referred to today (Henry & Mark, 2003), but

other types of uses have been identified by others. Patton (1998) complemented the evaluation use terminology with "process use". The identified types of uses can all be summarised as evaluation influence and are used in this thesis as analytical categories.

Instrumental use is probably the most dominant type of use in the research on evaluation and in talks about evaluation. It refers to use as an immediate and direct consequence or outcome of evaluation to modify the object of evaluation (Shulha & Cousins, 1997). Dahler-Larsen (1998) claims that many people seem to consider instrumental use as the ideal use. Conceptual evaluation is a type of use that helps local stakeholders, (e.g., university staff) gain new understanding, insights and ideas about the evaluand (e.g., a programme) based upon an evaluation (Weiss, 1998). Conceptual use has also been referred to as conceptual enlightenment (Fleischer & Christie, 2009). Symbolic use is understood as primarily symbolic, often employed to legitimise evaluation being carried out. Process use is a type of use that Patton (2008, p. 155) described as "a result of learning that occurs during the evaluation process", both for individuals and for organisations.

In this subsection I present key aspects that are known from research to increase use based on three literature reviews (Cousins & Leithwood, 1986; Johnson et al., 2009; Shulha & Cousins, 1997). Johnson et al. (2009) present the following definition of evaluation use: "The application of evaluation processes, products and findings to produce an effect" (Johnson et al., 2009, p. 378). They found two new aspects in addition to the ones that have been presented in prior reviews on evaluation use. The aspects earlier literature reviews have pointed to needed to be in place in order to achieve the intended use, relating to implementation and decision and policy setting. Examples of characteristics of implementations that were regarded as important to increasing evaluation use were: relevance and quality of the evaluation, communication quality and timeliness. Examples of important decisions and policy-setting characteristics for higher evaluation use include commitment or receptiveness to evaluation and an inhibitor to use is competing information about evaluation (Cousins & Leithwood, 1986; Shulha & Cousins, 1997). Support throughout implementation is an important factor if organisations want to achieve the intended purposes of evaluations (Cousins & Leithwood, 1986; Shulha & Cousins, 1997). Furthermore, authors of another review study suggest that evaluators who want to increase use should involve stakeholders, build evaluation "competence of individual evaluators, both professionally and culturally" (Johnson et al., 2009, p. 388). A guiding principle in UFE is to involve stakeholders actively by implementing evaluation at all stages of

a project or program. The involvement of the intended users is essential to ensuring that it is not evaluators alone who decide which evaluation models to use (Patton, 2008, 2018).

The focus at *intended use of evaluation* should be understood in relation to the tradition evaluation is rooted in. Højlund (2014, p. 28) states that "Most models of evaluation and evaluation use are derived from this primary rationale and function along logics of cause and effect. This is ontologically and epistemologically linked to realist and positivist understandings". Albæk (1995) argues that the positivist assumptions behind evaluation have been weakened as it was challenged by critical theory, hermeneutic and phenomenological traditions. Within the field of evaluation, participatory (Cousins & Whitmore, 1998), collaborative (O'Sullivan, 2004) and utilization-focused (Patton, 1997, 2008) evaluation approaches are examples of theories and beliefs in which the stakeholders are at the centre and in which evaluation is understood as processes and not only end-result findings. The three abovementioned approaches to evaluation are central to fourth-generation evaluations that consider evaluation to be a socially constructed practice by the involved actors. However, the cause-and-effect logics in positivism still dominate many evaluation models and practices (Sanderson, 2000).

The core concept in Utilization-focused evaluation (UFE) is to judge evaluation on its intended use for the intended users (Patton, 2008). By using UFE as an analytical approach, I defined intended users as academics and students, and intended use as improved teaching and learning (Borch et al., 2020). Experiences of intended uses described by the informants are central to this thesis, and also when answering the overall research question and exploring pedagogical dimensions of evaluation practice. However, as mentioned earlier, evaluation potentially has many intended users, inside and outside the university. An example of other users are actors who write and read documents describing and reporting on student evaluation practice, administrative staff, university management, educational leaders at different levels and in some cases NOKUT.

By including documents that are part of evaluation processes in this study, I am also exploring how these documents are used by actors other than my informants. This can be regarded as indirect use in macro-organisational processes (Mark, 2009), based upon my interpretation and analysis of the documents. In order to describe and discuss documentation routines and documents' role in evaluation practice, I am using terminology from evaluation use literature

and research on evaluation use. The types of evaluation use described above will be used as analytical categories in the thesis to illustrate multiple influences of student evaluation practice at the university. However, it is important to underscore that these terms are not used by the informants or written in the documents. This categorisation of student evaluation practice is my analytical interpretation of how evaluation is described by the informants and in the documents. Mark and Henry (Mark & Henry, 2004) have explored evaluation use and suggest that evaluation use or evaluation "consequences" can be distinguished at three levels of analysis: individual, interpersonal and collective. They suggest that when consequences of evaluations involve changes for a person, these changes are on an *individual* level, changes that affect interaction between individuals are on *interpersonal* level and the *collective* level refers to changes that affect organisations at a macro level. Changes at one level often affect another level (Mark, 2009; Mark & Henry, 2004); for example, a teacher's way of thinking at the individual level leads to change at the interpersonal level if the teacher discusses the findings with colleagues. By exploring student evaluation and how these evaluations are used, this study mainly analyses evaluation use from the informants' perspectives at the personal and interpersonal levels; however, the document analysis also expresses or describes organisational dimensions with evaluation use at a collective level. As change on one level can affect another level, and as evaluation often means something different to different people (King & Stevahn, 2013), it is neither possible nor my intention to identify all kinds of uses at the university. I will use different terms for evaluation use pragmatically in the analysis of the empirical data and in the discussion section of the thesis and the papers.

3.3 Translation theory

In order to better understand how evaluation is carried out and used, I rely on an understanding of ideas derived from new institutionalism and translation theory, and regard student evaluation as an idea that has already been translated into the university. I am exploring how evaluation is contextualised at the university.

I draw upon translation theory as described by Røvik (2011, p. 642). He defines translation as the "more or less deliberate transformation of practices and/or ideas that happens when various actors try to transfer and implement them". He has characterised translation as an "alternative doctrine for implementation" (My translation from Norwegian, Røvik, 2014, p. 403). In translation theory ideas are continuously changed *during* the translation and implementation

processes and are highly dependent on the actors responsible for the translation, how well they know the origin of the ideas and the contexts (arenas) these ideas are translated into (Røvik, 2014). This understanding of actors differs from the dominant hierarchal top-down doctrine in implementation theory where actors are viewed as passive receivers and ideas are understood as fixed, physical phenomena that are implemented like technical rational instalments and not dynamic processes (ibid). More than one translator is often involved in translation of organisational ideas, either at the same time or in different stages in the translation of practices, like chains of translations. Different actors influence the translation process of organisational ideas and their translations are based on their knowledge about the idea, the context it originates from and the one they translate it into (Røvik, 2014). In this study I have explored how the idea of student evaluation is contextualised at education programmes—meaning what the leaders base their translation on when they carry out evaluation at programme level.

In the second paper I present translation theory as an analytical framework to analyse and interpret how student evaluation was translated by the actors at the university (Borch, 2020). I regard student evaluation as a phenomenon or idea that is contextualised and translated at different levels and by different actors at the university. Translation theory "has proved to be a versatile analytical lens" used by academics with different disciplinary backgrounds (Wæraas & Nielsen, 2016, p. 236), and is a growing perspective among scholars within organisational research. It has been used by scholars in organisation and management disciplines who derive their work from institutional theory, actor-network theory and theories of knowing and learning (ibid). It is not my intention to present a comprehensive review and distinguish between features of the different perspectives but to illuminate that multiple versions of translation theory exist and are used by scholars with different disciplinary backgrounds[19] and going forward in this section to present the understanding of translation theory that I draw on in the thesis.

Røvik (2016) grounds his understanding of translation theory in actor-network theory and Scandinavian new institutionalism but also with inspiration from what he considers a neighbouring discipline: Translation studies. Whereas translation studies originate in linguistics, translation theory originates in new institutionalism. New institutionalism can be regarded as a way of thinking about social life and is grounded epistemologically in social constructivism (Eriksson-Zetterquist, 2009). Constructivist versions of institutional theory are

---

[19] For a more exhaustive literature review and presentation of what features and relationships there are between the different perspectives of translation in organisation research, see (Røvik, 2016; Wæraas & Nielsen, 2016).

developed by empirical studies concerning *how* institutions change and emerge—not only *that* they change and emerge (Czarniawska, 2008). Scholars within Scandinavian institutionalism are interested in understanding how ideas or phenomena are constructed and translated by active actors rather than adopted by passive receivers (Werr & Walgenbach, 2019).

Røvik uses several examples from management and management ideas in his dissemination of translation theory (Røvik, 1998, 2007, 2016, 2019). Student evaluation can be regarded as an idea that has to be translated into and within a local context—the university. Several organisational researchers have studied how different types of ideas from one origin are translated into local contexts. One frequently cited paper about ideas within institutionalism is "Travel of ideas" by Czarniawska and Joerges, they express that (1996, p. 25) "The perceived attributes of an idea, the perceived characteristics of a problem and the match between them are all created, negotiated or imposed during the collective translation process".

Although translation can be regarded as an alternative doctrine to implementation, it differs from implementation and other "knowledge-for-action theories" like knowledge utilisation, diffusion and transfer in many ways[20]. These theories have in common that they all "focus on knowledge and change" (Ottoson, 2009, p. 8). Ottoson (2009, p. 17) discusses these theoretical approaches and their possible implications on evaluation—she considers these approaches to be "additional theoretical lenses" that "complement evaluation theory by providing depth and perspective on the change process".

Røvik describes different modes of translation and different stages of translation processes. These are presented in paper two and will be elaborated upon below. "Translation processes normally involve translation of something from a source context (de-contextualization) to a target domain (contextualization)" (Røvik, 2016, p. 3). As this thesis is a study on student evaluation, a phenomenon that already has travelled into the university, it is mainly the contextualisation that is explored in the thesis. I have first-hand knowledge from the translators of evaluation on the programme level and not from those who developed and communicated the local quality assurance system. The modes of translation are reproducing mode, modified mode and radical mode. Each mode has different ambitions and translation rules. The reproducing mode aims to imitate a practice from another organisation as precisely as possible

---

[20] For a more comprehensive presentation of the relation between knowledge utilisation, diffusion, implementation, transfer and translation, see e.g., Ottoson, (2009).

with the translation rule of copying. Modified mode encompasses a pragmatic approach to translation in that it aims to recreate practices from other contexts to obtain the wanted result; however, the translator acknowledges that the translation must be adjusted to the specific organisational context. The translation rules are addition and subtraction of elements in order to better customise the idea to the organisational context where the idea is carried out. In the third mode, the radical mode, the ambition is not to copy other practices but rather to be inspired to develop one's own approaches, and the translation rule is transformation (Røvik, 2014).

Educational policy researchers have also used terminology from translation theory about translations being made and actors as active translators of ideas (Dahler-Larsen & Krogstrup, 2001; Lillejord et al., 2018; Stensaker, 2006; Steyn et al., 2019; Westerheijden & Kohoutek, 2014). Not many empirical translation studies are found in my literature reviews, however, and those I found were unsurprisingly from Scandinavia (e.g. Furu et al., 2020; Glosvik, 2017; Johannessen, 2020; Kalpazidou Schmidt, 2017).

3.4 Textual agency

Texts are undoubtably central to modern universities. Documents can be understood as artifacts or non-human actors. It is acknowledged that these texts are highly dependent on human actors in different ways (Atkinson & Coffey, 2011; Cooren, 2004; Prior, 2003). They are written and read by people, acted upon, given certain meaning or ignored by people. At universities, stakeholders like students, academics, management and administrative staff are spending a lot of time dealing with documents and systems that generate these texts. Some would even argue that universities exist through documents ("a university is in its documents rather than in its buildings" according to Prior (2003, p. 60)), and that documents "are essential to the objectification of organizations and institutions and to how they exist as such" (Smith, 2001, p. 159). Textual agency, as described by Cooren (2004, 2015), emphasises that documents are active contributors in organisational processes and constitute organisations together with human actors. Documents are considered contributors to organisational processes in that they can 'perform something'. They are durable and can influence people, for instance by reminding, confirming, indicating, informing, suggesting or predicting something. When ascribing documents agency, not only human agents in an organisation, but also their "nonhuman counterparts" (Cooren, 2004, p. 380) have the ability to perform something. When I in this thesis view documents as non-human actors that form organisational processes, I am

particularly interested in how student evaluations feature in documents, how student evaluation practices are documented and reported and how these descriptions may contribute to constituting understandings of student evaluation practice, though always in interaction with and interpreted by humans. Textual agency is chosen as the analytical framework in the third paper but the agency texts have or do not have is also relevant in the second paper and for the overarching research question. Textual agency is grounded in constructivism and is presented in paper three thoroughly, moreover in 3.3.2 Document analysis of the third paper (Borch et al., 2021).

4.0 Methodology and methods

4.1 Methodological approach

Drawing upon social constructivism, I will characterise my own role as a researcher as an active part of the society I am studying—my prior knowledge affects the research process in different ways. This way of understanding the researchers' role is central to social constructivism. In section 4.8.1 Doing research in one's own organisation and reflexivity, I elaborate on my role as a researcher and how it might have affected the research process.

I acknowledge that there are many different ways to investigate student evaluation practice. In this project student evaluation is explored from different perspectives. I aim to understand how student evaluation is carried out and used at the university from the perspectives of students and academics, and also how it is documented in evaluation documents. Thus, this project aims to include multiple perspectives of the phenomenon of student evaluation. Nevertheless, by investigating the phenomenon of student evaluation practice, wherein processes and relations are central, this is opened up for human interpretations by the informants and the researcher. Therefore, it is particularly important to be transparent about the research process.

4.2 Study design

This study has a qualitative exploratory data-driven research design. In order to answer the research questions and get rich and broad data about student evaluation practice from different perspectives, a qualitative research approach with a combination of methods was chosen: (1) Semi-structured interviews with academic leaders, (2) Focus group discussions with students

and (3) Document analysis of documents describing evaluation practice at different levels at the university. Each of the methods will be described in the sub-sections below.

In this study I use the terms "combination of methods" and "triangulation" where the former refers to something slightly different from the latter. The purpose of combining methods in this study was to get a comprehensive understanding of the phenomenon of student evaluation from different perspectives. When the term "data triangulation" is used in quantitative research it is often to validate the data by checking for bias; in qualitative research, particularly from a constructivist perspective, the purpose is to "capture the richness and diversity of perspectives on a phenomenon" (Varpio et al., 2017, p. 44). However, qualitative researchers are using the word "triangulation" to illustrate that different methods complement each other. In the case of this study, they complement each other to provide rich data about the phenomenon being studied from different perspectives. Malterud (2001, p. 487) stated that the aim of triangulation "is to increase the understanding of complex phenomena". In this study the multimethod approach helped answer the overall research question in that the methods complemented each other and together provided more insight about evaluation practice than one single method alone could. By combining different methods, I gained knowledge about different stages in evaluation processes from a student and academic perspective, and through document analysis I developed insight about the documentation and reporting stages of evaluation practice.

4.2.1 Local and national contexts

The local and national contexts, including educational policies, are described in the papers, particularly in the second paper. In this section, I will shortly present the national and local policies regulating student evaluation practice and provide additional information about the local context.

The Act relating to Universities and University Colleges (Universitets- og Høgskoleloven, 2005) states that the overall aims with local quality assurance systems (QAS) including student evaluations are educational quality assurance and enhancement. At the university, the local quality assurance system articulates that results from internal evaluations are foundational for educational quality assurance and enhancement of the programmes and further, that student evaluation is part of student empowerment and students' learning processes (UiT, 2012). The QAS was approved by NOKUT in 2012; the committee had some remarks regarding the system and advised UiT to improve some dimensions of the student evaluation practice. They

recommended that UiT conduct more evaluations *during* courses instead of at the *end* of courses, develop more course-specific questions and involve students actively throughout the evaluation process in order to improve the documentation of other types of feedback from students besides written evaluations and to strengthen the communication about evaluation results to the students (NOKUT, 2012).

UiT is the biggest regional higher education institution in Northern Norway, with 17 000 students and 3 600 employees. The study is carried out at one of the university's eight faculties, the faculty of health sciences. The faculty today has a total of 16 undergraduate and 20 graduate programmes whereof thirteen are health profession education programmes. Additionally, the faculty has a PhD programme, some annual course units and further education courses for health professions. The faculty has 4 700 students and 1 000 employees and a total of ten departments. The university has during the last two decades been through three mergers with university colleges[21] in the region and today has campus sites spread over the geographical county of Northern Norway. As many as nine of the health profession education programmes at the faculty have background from university colleges. Eight health profession education programmes are included in the study. The programmes in the study are located at the biggest university campus. Evaluation practice at these programmes is regulated by the quality assurance system from 2012 (UiT), the system is presented in paper two (Borch, 2020), moreover by internal procedures at each faculty (UiTø, 2010). After the university mergers, there was an agreement to accept the traditional educational programme leadership models from the prior university colleges and the university resulting in a complexity of different educational leadership models.

This PhD project was included in the research project SLUSEN[22] funded by the Research Council of Norway (Norges Forskningsråd, (NFR)) in 2017.

---

[21] In 2009 the University of Tromsø merged with the university college of Tromsø, in 2013 with the university college of Finnmark and in 2016 with the University colleges of Narvik and Harstad.

[22] Researchers at the Centre for Teaching, Learning and Technology, UiT applied for project funding before I was employed at the centre. Because the topic of the PhD project resonates with the one in SLUSEN, and the structure and participants in SLUSEN changed after the application, the project leader established dialogue with NFR and sent an updated project description of SLUSEN wherein this PhD project was included. For more information about SLUSEN, see https://prosjektbanken.forskningsradet and the project number 260359.

4.2.2 Inclusion of programmes, informants and documents

Informants and documents from eight of thirteen (A-H) health profession education programmes are included in the study (Table 1 below).

Table 1 Overview of empirical data

Programme	Focus group	Semi-structured interview	Document analysis
A	x		x
B	x	x	x
C	x	x	x
D		x	x
E	x	x	x
F	x	x	x
G	x	x	x
H	x	x	x

The inclusion of programmes was based upon a purposeful sampling strategy, which means that "the inquirer selects individuals and sites for study because they can purposefully inform an understanding of the research problem and central phenomenon in the study (Creswell, 2013, p. 156). I chose to include study programmes from a faculty I knew well where I had been employed prior to the PhD project[23]. Because of my knowledge about the faculty, I knew the programme portfolio was diverse. I aimed to include both undergraduate and graduate programmes from different departments at the faculty that differed in size and years of history at the university to ensure breadth and diversity.

The inclusion criteria for the academics were teaching experience, responsibility for planning, implementation and analysis of student course evaluations. I knew many academics at the university who could be included in the study based upon the criteria because the research is conducted in the same organisation where I am employed. However, I decided not to include former colleagues or teachers who I had collaborated with, except one colleague in a pilot interview. I knew this colleague well and believed she could provide constructive feedback

---

[23] I elaborate on advantages and disadvantages with doing research in one's own organisation in Chapter 4.8.1.

about my role as a researcher, the interview guide and the interview setting. The inclusion criteria were aligned with a purposeful sampling strategy (Creswell, 2013; Patton, 2015). The academics were contacted by e-mail. The e-mail contained information about the project and a request to participate (Appendix 6). Administrative staff helped me to identify potential informants based upon their knowledge about who fulfilled the inclusion criteria in the programmes I did not know well.

From a very early stage of the PhD project, I considered it important to include the students themselves and to give them the opportunity to share their experiences with student evaluation practices. Firstly, because they are the target group and recipients of the student evaluations, secondly because early literature reviews showed that there are few studies that explore student evaluation from their perspective. The only inclusion criterion for the students was that they needed to have participated in student evaluation. The inclusion criteria for the focus group discussions aimed to have student representation from different years of study, except first year students if they had not participated in student evaluations yet. An information letter about the project and a request to participate was sent to the student representatives of the included programmes and to students who represented the programme in department or faculty board by e-mail (Appendix 6). This can be described as purposeful sampling with the aim to include informants who could provide in-depth and rich information about the studied phenomenon (Creswell, 2013; Patton, 2015). The student representatives were invited to participate themselves or asked to help recruit students at the study programme they represented if they could not or did not want to participate themselves. I recognise that the included student informants may not be representative of the whole student group, as I recruited student representatives and student peers they recruited. They may be among the most engaged and active students at UiT.

The empirical documentary sources were collected from September 2016-May 2019. The first documents that I included were samples of student evaluation surveys and description of dialogue-based evaluation methods, educational quality reports from programme, department and faculty levels from 2013, 2014 and 2015. These documents served as a background before the interviews and some are categorised as supplementary empirical material in the first paper (Borch et al., 2020). Later in the research process, while working on the second and third paper, I decided to include more documents to get a better overview of the evaluation *system*. Below in Table 2, I provide an overview of the selected and analysed documents.

Table 2 Included documents

Type of document	Time period	Created by
Educational quality report Faculty level	2013-2015	Administrative staff on behalf of the faculty
Educational quality report Departmental level	2013-2015	Administrative employees on behalf of the department
Educational quality reports Programme level	2013-2015	Academic leaders on behalf of the programmes
Summaries of evaluation findings	2013-2015	Generated from the feedback platform Questback by administrative staff
Templates of student evaluation	2013-2015 and project period	Academics and administrative staff
Quality assurance system	2012 version, first approved by the university board in 2009	Administrative staff on behalf of the university management
Orientation letters of Quality assurance systems and revisions	2009, 2011, 2013	Administrative staff on behalf of the university management
Request letters educational quality from institutional management and faculty level	2013-2016	Administrative employees on behalf of the university management
Evaluation procedures at the Faculty of Health Sciences	Approved by the faculty board in 2010	Faculty management (Administrative and Academic leadership)
Board minutes from programme committees and department boards where educational quality reports or evaluation was on the agenda	2013-2016	Administrative staff

As the table above shows, the types of documents are created at different institutional levels, some are written by academics, other by administrative staff.

I planned to collect documents describing evaluation practices at different levels in the organisation—programme, department, faculty and institutional level. During the collection of data, I became aware of that not all programmes create educational quality reports. I had to choose, whether to include other documents or to continue with the plan and taking a different approach to the documents. I decided the latter. After all, it was a finding that reports were not

created at all programmes. Institutional educational quality reports from 2003 until 2018 were read but not included as they provided limited information about student evaluation practice.

Access to the documents was probably made easier because I was employed at the university and knew administrative staff with archive access, see also section 4.8.1. Doing research in one's own organisation and reflexivity.

4.2.3 Research strategies

The research process can be described as iterative or cyclic rather than linear, with an interplay between data collection and analysis. I started out with an inductive research strategy where I planned to conduct the focus group interviews and the semi-structured interview from the same programme, analyse each of them separately and then in conjunction with each other before I moved on to the next interviews with informants from a second, third, fourth programme, etc. The planned order of the interviews, focus group and analysis was not maintained throughout the data collection period and only for the first three programmes. After that, the informants' schedules became decisive as to when the interviews took place. During the preparation for the interviews, I established a preliminary overview of evaluation practice at the university by reading documents that regulate evaluation practice and educational quality reports which describe educational quality and student evaluation practice. This early document analysis turned out to be useful as background for both focus group discussions and the semi-structured interviews as it provided an extensive body of information about evaluation practice.

The documents played different roles throughout the research process. I returned to the documents several times between the interviews, during the analysis of the interviews and finally in the third paper which is based on documentary material and interviews as supplementary data. The research process can be described as data-driven but was not strictly inductive. The research process became an iterative process in which the different stages overlapped, as I moved between data collection, analysis and reading of theory. Thus, it can better be described as an inductive-abductive research process. Abductive and inductive research approaches are both data-driven, though inductive analyses are purely data-driven while abductive analyses also acknowledge that researchers have some theoretical knowledge that may be used to influence the analysis in interpretation and understanding of the empirical data (Alvesson & Sköldberg, 2017). According to Norman Blaikie (2007), a research strategy

is abductive when the researcher (like me) works "bottom-up" and derives concepts and theory from the situation. By having some knowledge about evaluation practice, theory helped me describe and analyse the empirical data i.e., with terms and concepts derived from evaluation theory. In abductive analysis the theory helps the researcher interpret and analyse the empirical data but the data also influence the researcher in choosing theoretical approaches (Järvinen & Mik-Meyer, 2017). The analysis of the empirical interview data will be described further in Chapter 4.5 and the analysis of the documentary material in Chapter 4.6.

4.3 Interviews

All interviews—the focus groups as well as the semi-structured—were conducted during the first year of the PhD project. They were based on interview guides with the same topics using the same analytical approach (Appendix 4 & 5). Nonetheless, my role as a researcher, the context, the inclusion strategies, the interview settings and aims of the interviews were different in the semi-structured interviews and the focus group discussions and will be presented separately below.

4.3.1. Semi-structured interviews

As already mentioned, I arranged a pilot interview as preparation for the interviews with the leaders. Before the pilot interview, I had created an interview guide consisting of six topics (Appendix 4 & 5). Each of the topics was followed by key questions and possible follow-up questions. The time frame for the pilot interview was rather open. I had in agreement with my colleague included time for a discussion and feedback about the interview, particularly after the interview but also during the interview if he experienced lack of clarity with the interview questions. The pilot interview lasted for 75 minutes and gave an indication of how much time I should estimate to the informants. I took notes on the interview guide during the interview. During and after the interview my colleague provided feedback about the interview that was used in further development of the interview guide together with my own reflections and notes.

A semi-structured interview is described by Brinkmann and Kvale (2015, p. 6) as "an interview with the purpose of obtaining descriptions of the life world of the interviewee in order to interpret meaning of the described phenomena". I regard the interviews as an interplay or interaction between the informants and the researcher. This way of understanding the role as a

researcher and the interview as an active process has been elaborated on by Brinkmann and Kvale (2015, p. 21) who consider interviewing to be a "social production of knowledge" in the relationship between the interviewer and the interviewee.

Seven semi-structured in-depth interviews with leaders were conducted between October 2016-June 2017. The interviews took place either in one of the university's meeting rooms or in the informant's office based upon what the informants preferred.

Each of the interviews started with a repetition of the study's aim, which the informants had already received by e-mail in the information letter with informed consent (more about informed consent in Chapter 4.8 Ethical consideration). If the informants had not signed the informed consent form before the interview, they signed it immediately before the interview began.

As described above, the interviews were based upon an interview guide. The interview guide was created by me in collaboration with my supervisors. It was based upon the aims of the study and the research questions and adjusted after the pilot interview. The interview guide was not followed closely but was helpful for me in order to prepare for and structure the interview. In the beginning of each of the interviews, I asked questions about the informant's position and their role in student evaluation practice. The interview guide included six main topics that described the student evaluation process, how evaluation practice was carried out and used, all followed by suggested sub-questions or follow-up questions. All the questions were open-ended and aimed to promote an interaction between the informant and myself.

The interviews were audio recorded and lasted between 75 and 90 minutes. During the interviews all topics in the interview guide were elaborated upon by the informants, but not all the sub-questions were posed. The scheduled time frame was held. After each interview, I wrote a reflection text and a short summary for myself. These texts are considered early stages of the data analysis.

4.3.2 Focus group discussions

Focus group interview or discussions as research method dates back to the 1930s and was considered to be an alternative to individual interviews and an interview format that shifted the

attention from the interviewer to the respondents (Krueger & Casey, 2015a). I chose to do the interviews with the students in groups because focus group is a well-suited method to explore perceptions or thinking about ideas or phenomena for groups of people and also "to dilute the power imbalance between researcher and researched" (Barbour, 2005, p. 743) which I expected to be between me as a researcher and prior head of studies and them as students. Focus groups discussions are emphasised as useful in eliciting the student voice in educational research (Barbour, 2005; O'Neill et al., 2002). By framing the interviews as focus groups, I could include more students than by conducting individual interviews. In focus group interviews the aim is not to establish a consensus or agreement among the participants, but to get a range of perceptions and opinions on the studied phenomenon (Rabiee, 2004). When planning a focus group, it is recommended to include informants who have something in common. One reason is that people are more likely to share if they have something in common (ibid). In order to provide a safe environment and encourage the students to answer honestly and speak freely, I underscored that I was an independent researcher who had designed the project by myself and I ensured them confidentiality.

In total seven focus group interviews were conducted in a meeting room at the university. The time frame for the interview was estimated at 90 minutes. I welcomed all the informants at the meeting room, repeated the information from the information letter, informed them about how the interview data would be used in paper publications and asked those who had not already signed an informed consent to do so before the interview started. The interviews were audio-recorded. In addition to the students and myself, an observer was present during the discussions. It is recommended to include an observer in the focus group settings with the purpose of assisting the researcher with practicalities before, during and after the meeting and to intervene or help to involve all participants particularly if somebody is very quiet. The observer was a colleague with experience from focus group discussions. In the beginning of the focus group, the roles of the students, the observer and myself were clarified and everybody introduced themselves. The group size ranged between three and seven students. There is no magic number regarding a focus group size—some suggest minimum five, but others suggest smaller groups of three to four informants to ensure that everybody will get the opportunity to share their reflections regarding the topic of the focus group (Barbour, 2008; Kitzinger & Barbour, 1999). Homogenic groups have been suggested as positive for the group dynamics and to strengthen the association effect (Krueger & Casey, 2015b).

The dialogues among the students went easily; they were engaged in the topic and steered the discussions after I posed the overall questions from the interview guide. Throughout the focus groups I asked follow-up questions to ensure that I understood the informants correctly, but all informants participated in the discussions and little involvement from the observer was needed. After approximately 75 minutes, I informed them that we had 15 minutes left of the scheduled time and it was time to close the discussions. When I later announced that the interview formally was over, thanked the informants and turned off the voice recorder, the discussions continued. The students said they considered the research project important. The observer gave me feedback about how she had observed the dynamic in the group, and we spent some minutes after each focus group to reflecting upon the discussion, which can be considered preliminary interpretations.

One focus group discussion was cancelled due to illness among two of the informants (the group size for the scheduled interview was four). Because of upcoming exams, it was impossible to find a new date for a new interview with student informants for this programme. Another programme was therefore included in May/June 2017.

The first focus group discussion was originally planned as a pilot interview to test the interview guide, the time frame and the setting with observer and interviewer. The informants were engaged in the topic and suggested that I add one question about empowerment. After the interview, I adjusted the interview guide due to their suggestions and experiences during the interview. I also added another question because of what they said during the interview about evaluation fatigue. I transcribed the interview and acknowledged that their reflections about student evaluation provided insight and knowledge to me about the topic I considered valuable for the PhD project. I discussed this with the supervisors who suggested I include the pilot interview fully if the informants agreed. All informants agreed and signed the consent form that they had received together with the information letter before the interview.

4.4 Document analysis

There are many different ways to approach and analyse documents in research. Traditionally, the contents of the documents have been analysed and are often used as supplementary empirical data in research (Prior, 2008). Scholars have suggested that there is time to acknowledge that documents also do things with organisations (Atkinson & Coffey, 2011) and

that researchers also should investigate the function or role documents have in organisations and study them in the organisational setting they are part of (Prior, 2008) because information derived through content analysis often are limited and narrow (Peräkylä & Ruusuvuori, 2011).

In this study, the documents played different roles at different stages of the research process and in each of the papers. Before the inclusion of programmes and informants, I collected and read documents describing evaluation practice to establish an overview of evaluation practice at the faculty. When I had included the programmes, I collected evaluation templates from evaluation surveys, at course and programme. I conducted a qualitative content analysis (Boréus & Bergström, 2017) of the templates and coded the material with descriptive codes and categories. The results of the content analysis of the evaluation templates are presented and discussed in the first paper (Borch et al., 2020). While working on the second paper, and by combining interview data and documentary material, I took a different approach aligned with Prior (2003, p. 77) who suggests that researchers should examine how documents "available to human actors are woven into specific forms of translation". I analysed QAS and how evaluation is contextualised and translated at the university. After I had written the first two papers, I went back to the documents to analyse how student evaluations feature in the included documents. I had the informants' description of student evaluation practice, including documentation routines as background when I read the documents. This triggered my interest and curiosity in following the flow of information about student evaluation in different documents at different institutional levels. Inspired by Prior (2003, 2008) and Cooren (2004, 2015), I intended to analyse the *use* or *function* of these documents as part of evaluation processes and how they interact with each other—the intertextuality. I therefore chose textual agency by Cooren (2004, 2015) as an analytical perspective for the third paper. This perspective is presented in Chapter 3.4 and in the paper (Borch et al., 2021).

4.5 Analysis of the interview data

From an understanding of interviewing as knowledge construction, the interview and analysis must be understood as intertwined (Brinkmann & Kvale, 2015). Early phases of the analysis (or reflection upon the data) started during the interviews and the writing of the reflection text after the interview. This can be described as informal interpretation of the interview setting and the interaction during the interviews more than analysis of the interview content. My own background and the analytical approach also affected the analysis of the data, as described in Chapter 4.8.1 Doing research in one's own organisation and reflexivity. The analysis of the

interview data can like the research process be considered as data-driven and abductive-inductive. The interview data in this study are based upon students' and academics' understanding and perspectives on evaluation practice. However, as this study draw on social constructivism, I acknowledge that my own understanding of evaluation as processes was central in the development of the interview guide, the interviews and also in the analysis. The data analysis must be understood as an iterative and not a linear process, however major stages are identified and described. This is a common approach to data analysis in constructivist research projects (Justesen & Mik-Meyer, 2010).

The interview data were transcribed verbatim by me in the Computer Assisted Qualitative Data Software (CADQUAS) system NVivo, which helped me structuring the data. During the transcription, I took notes and created memos, highlighting some quotes that I considered to describe the phenomena in particular insightful and informative ways.

I did a thematic analysis of the interview data, with coding, categorisation, interpretation and representation (Creswell & Poth, 2018). There are many different approaches of thematic analysis, it is common to code the empirical data in early stages but different ways how to do structure the data (Braun & Clarke, 2020; Lindgren et al., 2020). Although the analysis of the data was iterative with overlapping stages, I am presenting three main stages of analysis inspired by Lindgren, Lundman and Graneheim (2020). After the early coding, they (ibid, 5) write that there are "different ways to continue and deepen the analysis from the codes". I chose a strategy of coding (described below) and then moved on abstracting and interpreting the manifested content and developed/created categories before I continued and formulated themes. In the presentation of the stages in the analysis, I will use the same terminology for codes as Saldaña (2013), he describes thematic analysis but he names the whole process as coding. Although the three main stages of analysis are presented, I alternated between the stages and read transcripts to check the raw data behind the codes, categories and themes several times. This way of moving back and forth between the codes and the raw data was important to develop in-depth understanding of the empirical data. The interviews with students and academics from the first three programmes were conducted and transcribed consecutively programme-wise. The transcript from one interview was read and reread and analysed in conjunction with the focus group data at the same programme to establish a comprehensive understanding of student evaluation practice at the chosen programme. The order of the

interviews at the other programmes was decided by the schedules of the informants but data from the same programme was also compared and analysed in conjunction with each other.

Qualitative research methodologists have different opinions about how much of the data corpus should be coded (Saldaña, 2013). I decided to not leave out data from the interviews, except some minor parts like the backgrounds of the informants (confidentiality) and data that was not related to the topic of the project at all.

During the first cycle coding, the empirical data were sorted by codes that described the data—these were "descriptive" and "process" codes (Saldaña, 2013). The descriptive codes helped me establish an early overview of evaluation practice. In this stage I created descriptive "free-standing codes" but the more data I had analysed, the more the number of codes grew, and I started to group and connect the codes that belonged together in a hierarchical system. The structures in NVivo enabled a hierarchical coding by which I could identify main codes and divide them into sub-nodes. This helped me structure the data.

In order to better understand the phenomena and create meaning, categories were developed from the initial codes in the second stage of the analysis. In this stage some of the codes merged, others were split, and sub-categories were developed. NVivo allowed me to code in thematic subtopics, which I found very useful as many of the experiences/perspectives relate to each other or describe phenomena that relate to each other. Saldaña (2013, p. 234) describes the second cycle as an advanced way of reorganising and re-analysing data coded through first cycle methods with the primary goal of developing "a sense of categorical, thematic, conceptual and/or theoretical organization from your array of first cycle codes".

Many of these descriptive codes were kept and served as background for the presentation of empirical findings in the first paper like purpose of evaluation, dialogue-based and written evaluation methods and student involvement but the majority of the categories in the second stage were less descriptive than the codes created in the first stage. The thematic analysis became more abductive in this stage because many of the categories were informed by theory, e.g., process use and learning-oriented evaluation. In the next stage of the analysis, I grouped the categories that related to each other and developed broader themes e.g., evaluation capacity, improvement-oriented evaluation and evaluation foundations.

After all the data relevant were analysed in organised in codes, categories and themes, I looked for patterns like similarities and differences between the programmes, within the same programme and from the different informant group's perspective. I analysed the interview data independently; however, I discussed the coding process and findings with my supervisors during the analysis.

4.6 Analysis of documentary material

The analysis of the documents had a more pragmatic approach than the interviews. The document analysis I conducted early in the research process established an overview of some aspects of evaluation practice that were useful in developing the research questions and to get a brief overview of what characterises student evaluation practice. However, in order to answer the research questions posed in the papers, different approaches were chosen for each of the papers. I will therefore present the early stages of document analysis in this sub-section and the last stages of the document analysis in the papers in separate subsections. I will emphasise that documents often serve a variety of purposes in the same research projects (Bowen, 2009) and, in the case of this Ph.D project, also within the papers.

Early-stage document analysis had a primary objective of getting more background information about student evaluation practice at the faculty of health sciences. At this stage, relevant documents were identified. Particularly relevant were templates of questionnaires, descriptions of dialogue-based evaluation methods, the local quality assurance system, the faculty's evaluation procedures, evaluation reports and the parts of educational quality reports describing student evaluation. This information was useful in the development of the interview guide, as preparation for the interviews and for specifying the research questions. I also identified the authors and addressees of the documents in this early stage. In the next stage, data from the reports that did not describe student evaluation were left out, e.g., chapters that were not considered relevant to answering my research questions. I conducted a qualitative content analysis (Boréus & Bergström, 2017) and relevant texts describing student evaluation were coded and categorised with descriptive codes in NVivo. During the early readings of the documents, I read all institutional educational quality reports written at the university from 2003 until today. However, these educational reports were considered too generic to answer my research questions and therefore only served as contextual background information and not

empirical data in the study. The content was not analysed to the same extent as the educational quality reports from lower organisational levels.

4.6.1. Document analysis second paper

While working on the second paper I delimited the corpus texts by identifying the documents relevant to the research question. I decided there was a need to include other new documents to supplement the interview data. The documents that were included at this stage were texts regarding quality assurance system (QAS): board minutes and approvals of new versions of QAS and information letters to the faculties. This directionality, when research questions lead to the relevant texts, predominates in medical education research (Kuper et al., 2013), and can be considered a pragmatic approach common in projects with a constructivist orientation (Justesen & Mik-Meyer, 2010). The texts were selected in order to shed light on the research question, and to gain insight and new knowledge about the phenomenon of evaluation use within the organisation. After the interviews were conducted, I went back to the documents and read documents describing the programme the informants represented and the reports from the department where the study programme was located to get a broader picture of how evaluation practice was described across the institutional levels. I compared the content of the reports with the interview data. It was while reading the documents after the interviews that I became aware of how the descriptions of student evaluation practice differed in the documents and in the interviews. This became the rationale of the third paper. As I used translation theory as an analytical approach in the second paper, and explored how evaluation was translated, I expected that the local quality assurance system would be one of the sources for translation and analysed the content of QAS. I categorised the different parts of QAS that comprise evaluation with descriptive codes similar to the thematic analysis of the interview data. I went back to these descriptive codes several times, during the analysis of the interview data and in the writing process of the second paper. This re-reading of the coded QAS helped me in the discussion of the empirical data and particularly in comparing how informants described translation of evaluation, evaluation practice and how the guidelines set by the university were formulated.

The content and the agency of the documents are analysed in the paper, i.e., both the content and the agency of the quality assurance system are analysed and discussed. The guidelines in QAS may also be described as intended evaluation practice, and the agency as the actual enactment of QAS or translation of evaluation. The interviews with the academics are essential

to understanding the translation of evaluation and how QAS is enacted by the academics at programme level. The interview data are the primary data source for the second paper and documents the secondary.

4.6.2 Document analysis third paper

The document analysis in the third paper is described thoroughly in the method section of the paper (Borch et al., 2021) and the analytical approach (textual agency) is presented in Chapter 3.4. I will present how the analysis differed from that in the first two papers in this subsection. I understand the documents as constructed and contextualised by people in the organisation, documents' role and function in evaluation practices is central to the second and third papers. However, different analytical approaches were chosen in the two papers. As mentioned earlier, higher education institutions are required to establish quality assurance systems that include student evaluation. They are also required to document and report on educational quality to the institutional board, and in case of auditing also to NOKUT (Studietilsynsforskriften, 2017). Therefore, the documents included in the third paper are part of a national policy context and some of them, like the educational quality reports, are part of a required documentation practice. Because the rationale for writing the third paper was based upon a divergent picture of student evaluation practice in the reports and in the interviews, I decided to analyse the documents from an analytical perspective that incorporated both the function and content of the documents in this paper. In the first stage analysis, I conducted a content analysis consisting of overlapping stages starting with superficial reading, followed by more thorough reading and interpretation (Bowen, 2009). The data were coded and categorised in NVivo in the next stage before I chose to analyse the data from a textual agency perspective (Cooren, 2004, 2015). By choosing textual agency as my analytical approach, I aimed to study the agency of the documents requesting (directives) and reporting (assertives) on student evaluation practice, as reporting is part of evaluation practice. In order to analyse the agency of the documents, I compared the description of student evaluation in these documents in relation to the interview, and what the academics said about documentation routines and reporting. The intertextuality between the documents and the relationship between how student evaluation practice is described by the academics in the interviews and in the documents from programme level and the description of student evaluation written by administrative staff at department and faculty level are discussed and analysed in the third paper and in the discussion section of the thesis. The interviews with the academics together with the documents made it possible to analyse the agency actors in

evaluation practice confer upon these documents. Analysing the function and use of documents together with the interviews contributed insight about evaluation practice different from what one of the methods alone could have gained and different from what pure content analysis of the documents could have given.

4.7 Methodological considerations

4.7.1 Quality in qualitative research

An overall aim when judging the quality in qualitative research is to achieve trustworthiness of a study, and central to achieving this purpose is a transparent documentation of the research process (Creswell & Poth, 2018; Hammersley, 2007). Rolfe (2006, p. 305) states: "A study is trustworthy if and only if the reader of the research report judges it to be so". It is important to be transparent about the researcher's background because qualitative research is grounded in an interpretative research paradigm where researchers' understanding of nature and being in the world (ontology) and the nature of knowledge (epistemology) affect the research approach and all stages of the research process (Alvesson & Sköldberg, 2018). I elaborate on my background and role in the research process in the subsection 4.8.1 Doing research in one's own organisation and reflexivity.

In this thesis I use different terms to argue for this study's trustworthiness, including validity and reliability; however, I emphasise that the latter terms are understood and defined differently in qualitative research than in quantitative. While internal validity refers to whether a study investigates what it intends to, "external validity asks in what contexts the findings can be applied" (Malterud, 2001, p. 484), and reliability in qualitative research comprises consistency (Cypress, 2017; Leung, 2015) and may be better referred to as dependability (Lincoln & Guba, 1985). When judging the quality of qualitative research, it is not only quality of single elements of the research process or the design that is of interest but rather also the interrelationship and consistency between methods, methodology and epistemology throughout the research process (Carter & Little, 2007). In the subsections of this chapter, I use the terms "validity as craftmanship and credibility", "dependability", "transferability" and "reflexivity" to argue for the study's trustworthiness.

4.7.2 Validity as quality of craftmanship and credibility

Maxwell (2013), among others, argues that validity is not something that can be taken for granted based on the methods, as it is rather a property of inference. Because qualitative researchers are part of the world we are studying and qualitative research uses emerging rather than predetermined approaches, different researchers are most likely not getting the exact same results (Denzin & Lincoln, 2018; Maxwell, 2013). Brinkmann and Kvale (2015, p. 283) consider validity as "quality of craftmanship". They describe seven validation stages but underscore that validity must permeate throughout the whole research process (ibid). In the following section I describe how I have strived for quality in this project throughout the research process.

The quality of craftmanship relates to the credibility of the researcher throughout the research process (Brinkmann & Kvale, 2015); more specifically, credibility means confidence in the truth of the data and the analysis (Polit & Beck, 2020), and credibility of the researcher's interpretations (Silverman, 2013). The validity of the study relates to how a researcher interprets the data, but also whether or not the inferences the researcher makes are supported by the data and by earlier research (ibid). I based the study design and development of the research questions on literature reviews that I conducted in early stages of the PhD project.

As mentioned above, the intention of exploring student evaluation from different perspectives was to get rich data about student evaluation practice. In order to accommodate this, I decided to combine different research methods. In some literature about qualitative inquiries this is described as triangulation of methods (e.g., Creswell & Poth, 2018). Varpio et al. (2017, p. 45) challenge the use of triangulation in qualitative research because combinations of methods in post-positivist orientation are ways of "enhancing the rigour through comprehensiveness rather than convergence". The intention in this study was not to find a consensus among the different perspectives and a shared understanding of student evaluation practice among the informants but rather to get a rich understanding of the studied phenomenon. However, my analysis of the empirical data showed that many of the findings from different methods converged and were similar, e.g., how students and academics described little or no direct student involvement in evaluation processes like development of evaluation questions (Borch et al., 2020). I would therefore refer to Barbour (2001, p. 1117), who, in the debate about combining different methods of qualitative research, said that: "(…) the production of similar findings from

different methods merely provides corroboration or reassurance (…)". Yin (2018) described combining data from multiple sources as corroboratory strategies for strengthening the quality of a research design.

Validity during the interviewing pertains to the trustworthiness of the informants' stories and the quality of the interviewing (Brinkmann & Kvale, 2015). In the interview settings, I posed clarifying questions if I was not sure I understood the informants correctly and I asked follow-up questions when I wanted them to elaborate and explain statements or expressions. Further was the credibility strengthened during the analysis by, for example, discussing the focus group findings with my research colleague and the observer in the focus groups. Immediately after each focus group we discussed the interactions within the groups and the overall expression of what characterised evaluation practice at the programmes.

Valid transcription of the interviews from oral to written format was ensured in that I transcribed the interviews verbatim. The quality of the recordings was of high quality. Some methodology literature suggests doing member checking of the analysis before publishing the results. This relates to what has been described as construct validity, meaning that the categories the researcher uses are meaningful to the informants and reflect the participants' experiences (Cohen et al., 2017). I did not send the transcribed data or paper drafts to the informants but shared the published papers with the teacher/leader informants and received feedback that they recognised evaluation practices described in the papers. I also discussed the analysis of the interviews with my supervisors and co-authors. This can be understood as a way to internally validate the data analysis, to establish a shared understanding of the data and enhancing the credibility of the study (Brinkmann & Kvale, 2015; Creswell & Poth, 2018).

Validity in document analysis was strengthened in this study by presenting and discussing the documents with my supervisors early in the research process. The interpretation of the documents' content, meaning and function were validated through discussions with my supervisors and co-authors during the writing of the papers.

4.7.3 Dependability

Dependability refers to how research findings may be affected by contextual relationships and changes (Lincoln & Guba, 1985). The term is often used as a synonym of reliability. They relate

to each other, but as qualitative research has different aims from quantitative and does not aim to statistically generalise, the criteria for judging the "reliability" or dependability must also be different for qualitative than for quantitative studies. Some may argue that it is inappropriate to use the term reliability in qualitative research, as the aim is not to replicate the exact same study (Blaikie & Priest, 2019).

Despite the fact that several aspects of the different interviews are the same in the present study, such as having the same interviewer, interview guide, same information letter, etc., each interview is unique in that the context, different researcher and informants will influence each other (Blaikie, 2007). Brinkmann and Kvale (2015, p. 260) use the term "interviewer reliability" and discuss this in relation to leading questions and the fact that "different interviewers get different results". I acknowledge that if the interviewer was a different person than me, the interactions would have been different.

As the documents were produced before this study, they were produced for a different purpose than research and can be approached in different ways. The analysis and results depend among other aspects on the theoretical approach, which in the case of this study is textual agency. If the documents, however, had been analysed from perspectives that focused on linguistic aspects or genres, other aspects and content of the documents most likely would have been elaborated upon.

4.7.4 Transferability or analytical generalisation

The intention of qualitative studies is not statistical generalisation and to be able to generalise, but rather to provide in-depth knowledge about phenomena. The term "analytical generalization" (Brinkmann & Kvale, 2015, p. 297) refers to "a reasoned judgement about the extent to which the findings of one study can be used as a guide to what might occur in another situation". When Yin (2018, p. 37) uses the term "analytical generalization" for case studies, he states that the role of theory is "the basis for analytical generalization". He suggests to "shed empirical lights on theoretical concepts or principles" in order to either modify, corroborate or rejecting existing theoretical concepts or develop new ones based upon the empirical findings (Yin, 2018, p. 38). I close the discussion section by proposing a framework for a more learning-centred and systemically aligned evaluation practice, a method of analytical generalisation in which my empirical findings are applied to an existing system alignment concept.

However, much of the responsibility for judging to what extent findings are transferable to other settings is left to readers, who must consider if the context of the study is similar to their own context (Kuper et al., 2008). The researcher must provide thick description of the context in which the research is carried out if the reader shall be able to make such judgment. The reader must also know the context to which the findings might apply well. Guba and Lincoln (1985) suggest using the term "transferability" to describe how results may be transferred from one context to other contexts in qualitative research. I aimed to provide thick contextual description to help the readers examine to what extent the findings might be transferable to their own, but it is still up to the reader to judge.

4.8 Ethical considerations

Research ethics conducted within Social Sciences and Humanities in Norway is regulated by ethical guidelines from the Norwegian National Research Ethics Committee (NESH), which comprise a complex set of values and standards regulating scientific activities[24]. This project is aligned with these guidelines. Additionally, this project is registered in a national archive for research data: the Norwegian Centre for Research Data (NSD[25]). I received a formal approval to collect and store interview data from NSD before the data collection started (appendix 1). During the project period an application for a project period extension was sent to NSD. I applied for a project extension because my position as a PhD candidate was put on hold while I was appointed as an assistant professor at the Centre for Teaching and Learning at UiT for six months during the project period. NSD approved the application (appendix 2).

Ethical considerations must be taken throughout the whole research process and on different levels. On a macro level, the researcher should reflect upon the study design and its value for society (Stutchbury & Fox, 2009). In the case of this research project, I consider the project to be relevant and valuable for society as we, according to Dahler-Larsen (2011), live in an "evaluation society" and collect a high number of evaluations in the public sector annually—evaluations that are time-consuming and cost an unknown amount of recourse without knowing much about how these evaluations are used. On a micro level, ethical considerations concern the individual participants in research and how they may be affected by their participation

---

[24] For more information about NESH guidelines: https://www.forskningsetikk.no/en/guidelines/.
[25] For more information about NSD, see webpage: https://www.nsd.no/en/.

(Stutchbury & Fox, 2009). Tangen (2014, p. 679) proposed an Ethical Matrix Model for educational researchers and categorised three ethical research domains based on the NESH guidelines: (1) Ethics *within* the research community; (2) protection of research participants; and (3) the role and value of educational research in society. How I approached the three domains and how ethical guidelines are followed throughout the research process is described below.

I will start by presenting how I strived to protect the participants. The inclusion criteria are described in the papers, and in section 4.2.2. All informants were contacted by e-mail with a request for research participation. The e-mail contained an information letter about the overall purpose of the project, a description of the research project, what participation involved and an informed consent. Informed consent entails several aims, from informing the participants about the purpose of the study and the features of design, that participation is voluntary and how the data will be used and stored (Brinkmann & Kvale, 2015). In the informed consent form of this study (appendix 6), the above-mentioned information was included. Furthermore, the informants were informed that they at any time of the study could withdraw from participation, that the research data would be treated with confidentiality and that interview data would be anonymised before presentation and stored on a password-protected computer and deleted three months after the project dissertation. The informants and the programmes are anonymised and presented with letters A-H. The information letter and content form were also included in the application to The National Ethical Committee (NSD) (appendix 1 & 2).

Returning to the three dimensions mentioned in the introduction of this section, I would like to add some reflections on how I have strived to approach the research data ethically within the organisation and how I regard the role and value of educational research in society. As part of my PhD position, I had some teaching obligations, which is quite common in Norway. During my project period I received several requests to share my knowledge about the topic, mostly from educational leaders within the university. In response I have run workshops, seminars, webinars and given lectures about student evaluation for academics, managers, student representatives and administrative staff. Early in the project period before I had analysed the findings and published my research, I was conscious not to present preliminary findings. I based my teaching on evaluation theories, research on student evaluation and the existing educational policy framework that were meant to direct evaluation practice internally. Later in the project period after I had published my work in academic journals, I included some of my findings in

the teaching. To my understanding this is aligned with ethical guidelines about how to handle research data within the organisation and the first dimension described by Tangen (2014) above.

Judging the research's value to society is primarily up to the readers. Nonetheless, I would like to add that I am aware of the fact that some employees within the sector may regard the topic as sensitive and the findings as negative, this because they illuminate challenges with today's evaluation practice at the included programmes. Therefore, express my gratitude for the honesty and insight the informants have provided because I believe their experiences contribute important knowledge about how evaluation practice could be improved. Another reflection regarding the positive contribution or value of the research to society and people outside the organisation is that it may enlighten the debate about student evaluation with empirical findings that may challenge some of the assumptions and myths often heard in the debate.

4.8.1 Doing research in one's own organisation and reflexivity

Being a researcher in familiar settings or one's own organisations has disadvantages and advantages, which Mercer (2007, p. 5) described as a "double-edged sword". Although easy access to the field can be described as an advantage, being familiar with the context or field of study might also be a disadvantage if the researcher is more likely to take things for granted or assume that their own perspectives are more widespread and thus develop myopia[26] (ibid).

Alvesson (2003, p. 174) provided self-ethnography as a frame of reference for researchers who are active participants in the cultural setting they want to study and use "experiences, knowledge and access to empirical material for research purposes". Although this is not the chosen research approach in this study, my background from the university and experiences with student evaluation have influenced the research process and my access to the field. Alvesson (2003, p. 187) also points at disadvantages and advantages of doing research in familiar settings or one's own organisation. He states that researchers have different context-dependent attitudes and identities; moreover, he considers having "multiple work-related social identities" an advantage because it makes it "easier to avoid being caught in a staying native position". This statement resonates with me and the importance of being transparent about my own background as I was employed at the university prior to this PhD project. In methodology

---

[26] In Merriam-Webster dictionary Myopia is defined as "a lack of foresight or discernment: a narrow view of something."

literature is this is often defined as *Reflexivity*. Reflexivity can be understood as deep and self-critical introspection and consciousness about one's perspective (Patton, 2015) or a person's ability to think "critically about our own assumption and actions" (Cunliffe, 2004, p. 408). Cunliffe (2016) elaborates about reflexivity in research and describes this as the researcher's ability to reflect upon his/her role in relation to the investigated phenomena. She emphasises that the researcher must be aware of how his/her understanding and background can have an impact on how he/she understands the world and the research process.

I consider it particularly important to reflect upon how my prior and current job positions might have affected the research process from the beginning to end because I did research in the organisation I am employed in. I aim to be reflexive and transparent about my own background in the thesis. Some of the informants, both academics and students, had met me prior to the invitation to participate in the project and when I held the position of head of studies. I therefore clarified for them before the interview that this project was an independent project. The relation between me and my role as a researcher and the students as informants in the interviews are asymmetric, and this asymmetry may be strengthened because I had worked as a head of studies together with leaders at the programmes they represented. Therefore, it was particularly important for me to inform the students about how the data would be treated in alignment with ethical guidelines for research and assure them that all informants would be anonymised in the dissemination of the data.

Being a researcher in my own organisation also gave me some advantages that can be considered a strength of the project as long as I am transparent about it. The advantage of getting easier access to the field is already mentioned in the method section 4.2.2. I would also argue that my background and prior experiences with student evaluation can be considered an advantage because this helped me in designing a research project of relevance to different stakeholders in the organisation. During my time as a head of studies I had developed a curiosity about many aspects related to evaluation practice, some of which were integrated in the interview guide. Many of the informants said to me that they regarded the project as important and pointed to the need for more knowledge about student evaluation in higher education.

5.0 Results

In this chapter, I am providing an overview of the main findings in each of the three papers. Figure 1 below illustrates how the papers relate to the overall research question and each other.
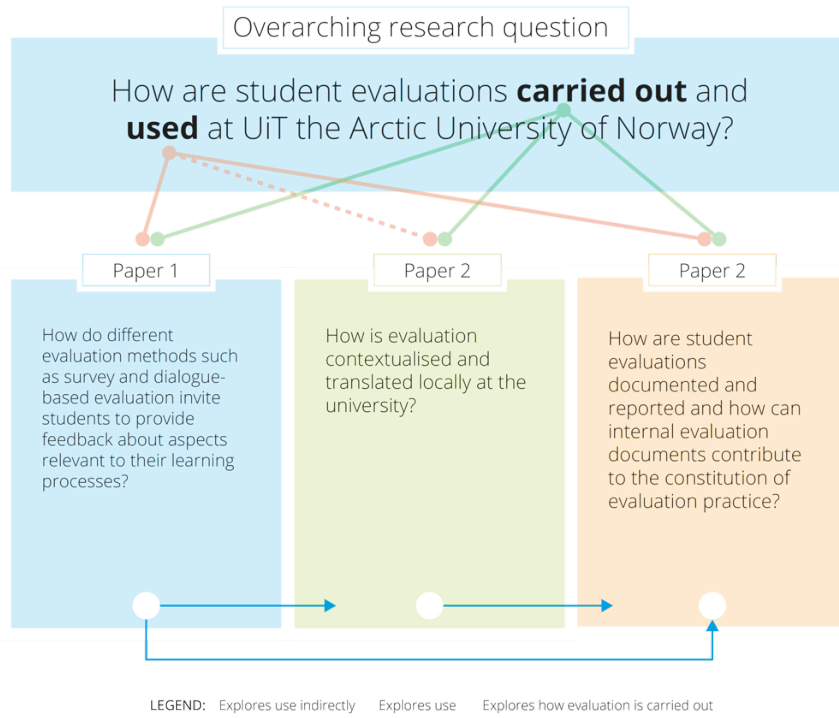

Figure 1: Relationship between papers and research questions

5.1 Paper 1

Borch, I., Sandvoll, R., & Risør, T. (2020). Discrepancies in Purposes of Student Course Evaluations: What Does It Mean to Be "Satisfied"? *Educational Assessment, Evaluation and Accountability*, *32*(1), 83-102.

The aim of this paper is to explore the use of student evaluation in relation to the overall purpose of educational evaluation: improved teaching and student learning. Drawing upon principles of Utilization Focused Evaluation (UFE), the paper investigates this intended use in relation to the evaluation methods. The research question is: How do different evaluation methods, such as surveys and dialogue-based evaluation, invite students to provide feedback about aspects relevant to their learning processes? In UFE, intended use should be judged by its utility for intended users. In this study the intended users are students and academics; students and academics are interviewed, and evaluation templates are analysed.

The main findings: The results of the study showed that the focus and the types of questions were different in written and dialogue-based evaluation. Written evaluation was described as teacher- and teaching-oriented, non-specific and superficial by the students. Many of the questions were also experienced by the teachers as non-specific and not suitable for teaching and course improvement that could facilitate students' learning processes. Dialogue-based evaluation, regardless of format, was described by the students as more meaningful, specific and learning-oriented. The dialogue-based evaluation had a more open format than the written surveys: while the surveys had predefined questions, the students could steer the discussions and focus of the dialogue-based methods. Students and teachers expressed that they benefited from the evaluative dialogues. The students said these dialogues helped them reflect upon the perceived learning, which in turn increased their awareness of the learning process. Some of the students expressed that they developed their communication skills. The teachers received valuable feedback from the students about perceived learning and how the learning activities had contributed to students' learning.

The paper's contribution to the field: The findings presented in the paper contribute knowledge about how evaluative dialogues about courses and teaching can be a viable supplement and/or alternative to written evaluation methods and be more actively used in pedagogical planning. It also illustrates that the key stakeholders—students—expect all evaluation methods to be learning-focused. If the students were more actively involved in planning of evaluation, aligned with research on evaluation use that suggests including key stakeholders, the learning perspective might be strengthened in all evaluation approaches. Today many of the questions in written evaluations are developed by administrative staff and are not specified to the courses or learning outcome descriptions. These implications for the field as to why and how to involve students more actively throughout the evaluation processes are discussed in the discussion section of the thesis.

5.2 Paper 2

Borch, I. (2020). Lost in Translation: From the University's Quality Assurance System to Student Evaluation Practice. *Nordic Journal of Studies in Educational Policy*, *6*(3), 231–244.

The aim of this paper is to explore how student evaluation is contextualised and translated locally at the university. Drawing upon translation theory, the paper analyses characteristics of the Quality Assurance System (QAS), the arenas where evaluation takes place and the actors (academic leaders) who are central to the planning and translation of student evaluation. The following research question was posed: How is student evaluation contextualised and translated at the university?

The main findings: The results of the study showed that student evaluation practice varied from programme to programme, and also within the same programme. The academics felt left to themselves in evaluation practice and said there was little communication about evaluation among their colleagues and with the leaders. They called for more support throughout the evaluation processes and better feedback routines. Some of them were uncertain whose responsibility it was to follow up on evaluation results and requested more feedback from their leaders about evaluation. The evaluation practices diverged from the evaluation guidelines and requirements in QAS and the actual evaluation practice was de-coupled from the intended practice described in QAS. Each academic teacher took a pragmatic approach and ensured student evaluation took place. The leaders based their translation on previous experiences, local cultures and tradition rather than QAS. Their translation of evaluation can be regarded as modified translation. QAS was developed and communicated by administrative staff on behalf of the university management who expected that the academics would act upon evaluation according to the guidelines and local requirements aligned with a logic of consequences. This was not the case, as the academics did not consider themselves knowledgeable about evaluation nor the details of QAS but the local contexts and the requirement to evaluate. They based their translations on their interpretations of the idea of student evaluation. Their translation was done within a logic of appropriateness. Contradicting logics may be an explanation as to why actual evaluation practices were de-coupled from the one described in QAS.

The paper's contribution to the field: The findings in this study showed that the local translation of evaluation should not be underestimated at the university. Although the university has a formal local quality assurance system with evaluation guidelines and each faculty their own procedures, the details of these are not well-known to the academics. Universities should make sure that they have evaluation plans, that their key stakeholders have a sense of ownership, and moreover that internal communication systems and practices that ensure that actors are familiar with their responsibilities and roles in evaluation processes.

The complexity of student evaluation in higher education can be augmented by institutional theory to get a nuanced understanding of how actors and contexts affect how a specific policy framework is translated into practice. There is great potential at the university to involve academic developers more actively in student evaluation processes, particularly to carry out contextualised evaluation practices with the intended purposes described in QAS. This will be discussed in the discussion section of the thesis.

5.3 Paper 3

Borch, I., Sandvoll, R., & Risør, T. (2021). Student Course Evaluation Documents: Constituting Evaluation Practice. *Assessment & Evaluation in Higher Education*.

The aim of the paper is to explore how documents can contribute to the constitution of evaluation practice. The following research question was posed: How are student evaluations documented and reported at the university and how can internal documents contribute to the constitution of evaluation practice?

Main findings: The findings showed that there was a difference between how students and academics experienced student evaluation practice at programme level and how student evaluation was described in the higher-level internal university documents. QAS states that the documents and evaluations from one organisational level shall be aggregated on the next one. Our analysis of the requested information as part of the documentation procedures showed that the questions posed from higher to lower levels at the university were mainly asking how evaluation was carried out and followed up and not about student evaluation results and perceived education quality. From a textual agency perspective, we analysed how these questions are appropriated by and attributed to actors, suggesting that this contributes to a certain picture of evaluation practice. The analysis, however, found misalignments between how student evaluation was documented at different organisational levels. Information about educational quality documented at programme level described student evaluation results, strengths and weaknesses with courses and programmes, challenges with student evaluation practices, including aspects that needed to improve and suggestions about evaluation results that had to be followed up, etc. The information in programme reports were similar to what the informants expressed in the interviews. Information from programme reports was nearly absent

from reports at department and faculty level. These educational quality reports concerned compliance between the evaluation guidelines stated in QAS and actual evaluation practice, and no description of weaknesses with programmes or evaluation practice. They described a well-established evaluation practice. Moreover, the student perspective about educational quality and evaluation practice was absent from these reports.

The paper's contribution to the field: Despite the high number of documents written as part of evaluation practice, there are very few published studies about documentation routines in evaluation. This study therefore contributes knowledge about a rather unexplored part of student evaluation in higher education. The findings in this study point to a need for increased awareness about several aspects of the documentation process. First, that questions as part of documentation procedures can be understood and interpreted differently by different stakeholders. Second, that documents may have a constituting effect on how evaluation is understood in the organisation. Third, when information described by academics that may contribute to improvement of evaluation practice is not aggregated to the next level, the documentation routines can be considered a ritual more than a practice that promotes organisational learning. When the reports at faculty level draw an image of evaluation as a successful activity in compliance with the system, this is in contrast to what the informants expressed in the interviews and from my analysis of the documents at programme level. If the reports at the faculty level are the only information about evaluation practice communicated to a higher institutional level, they give an impression of evaluation practice carried out in alignment with the university's intentions. This may hinder improvements of evaluation practice and may also contribute to an understanding of student evaluation as an accountability tool assuring educational quality rather than a practice that can be used to improve teaching and promote learning. Why this is unfortunate and how the focus in evaluation reporting may change in the future for a better alignment with the purposes of internal student evaluation will be discussed in the discussion section.

6.0 Discussion

The overarching research question in this study was: How are student evaluations carried out and used at UiT? With empirical data derived from interviews, focus group discussions and document analysis and through use of different analytical perspectives, an understanding emerged about student evaluation practice and how evaluations are used from different actors'

perspectives across organisational levels. The analytical perspectives are all drawing upon constructivism and emphasize that practices are constituted or constructed by actors in local contexts. I aim to discuss how different actors contribute to constituting evaluation practice within the university. Based upon my empirical data, I elaborate on different types of use in the first part of the discussion. To illustrate the types of use I will refer to the papers. Next, I will discuss how evaluation practices are carried out and by whom. I aim to discuss how pedagogical and organisational dimensions of evaluation practice seem to affect how student evaluation is carried out and used at the university, before I close the discussion by proposing a framework for a more systemic and student-learning-centred approach to student evaluation for the future.

6.1 Evaluation use in student evaluation practice

Central to the findings about student evaluation use at UiT is the gap between the intended use manifested by the university management in QAS and the actual use described by the informants. The evaluation practice can be regarded as decoupled from the evaluation system and the intention set by the university (Borch, 2020). The study also reveals that evaluation is used differently by actors at different institutional levels.

As a starting point of reference, I am posing the question: Why are we evaluating? Programmes, courses and teaching in higher education are evaluated for different reasons, which both this study and literature about student evaluation show. Based on the multiple purposes of student evaluation one could expect evaluation use to be diverse. I will in the following discuss different kind of uses that I found through my analysis of the empirical data. In short, student evaluation practice at UiT seems to be better at educational quality assurance than quality enhancement and decoupled from QAS. This study found possible explanations as to why evaluations are used the way they are throughout the evaluation process; in planning of evaluation, when evaluation is carried out and reported.

The first paper analyses pedagogical dimensions with evaluation practice. We point at possible limiting aspects of using evaluations in educational improvements to be found in the evaluation format, particularly the questions posed in surveys. These questions were teaching-focused and non-specific, asking about students' satisfaction rather than students' learning, contrary to the dialogue-based evaluations that invited students to dialogues about their learning processes (Borch et al., 2020). Darwin (2016), who also identified a similar focus in student evaluation

practice, argued that evaluation questions are consumerist-driven, asking what the students want instead of what they need to effectively learn. Such questions based on satisfaction ratings may be used in creating education that satisfies students, but this is not necessarily education that facilitates student learning. The students in the present study said they perceived dialogue-based evaluation to be valuable for their professional development and learning because the questions in the evaluation dialogue made them reflect upon their learning processes (Borch et al., 2020). This learning occurs *during* the evaluation process and is described as *process use* (Patton, 2007). The teachers expressed that they found dialogue-based evaluation and informal evaluation during the courses valuable for adjustments of the teaching approaches. By relating the teachers' descriptions to literature about evaluation use, I have categorised the types of use they referred to as *conceptual use* and *process use*, where conceptual use refers to new understanding and insights gained because of an evaluation (Weiss, 1998), and process use, as already mentioned, to learning that occurs during the evaluation process (Patton, 2007). These examples illustrate how evaluation can contribute to individual and interpersonal reflections about learning processes that in turn can improve learning and teaching, which is also a designated aim with student evaluation practice (Roxå et al., 2021; Ryan, 2015). The academics gave some but few examples of poor evaluation results that led to direct changes in courses, like replacing external supervisors with other clinicians in practical placements. This type of use can be considered *instrumental use* (Shulha & Cousins, 1997).

The second paper focuses on organisational dimensions of quality work and concludes that there are aspects of how the evaluation system is translated that affected how evaluation is carried out, hence indirectly used (Borch, 2020). The paper points at some hindrances to intended use found in the organisational context, such as the arena for translation and the translator's knowledge about evaluation. The leaders describe that they expected more incentives to use evaluation from their leaders. They were uncertain about whose responsibility it is to follow up on evaluation results, which can be a reason why some teachers are not using evaluations more actively. However, they knew evaluation was mandated. This may relate to the powerful driver of a law regulation (Scott, 2014), which most likely also affects communication about and institutionalisation of evaluation. The quality assurance aspect of evaluation is probably strengthened by reminders from administrative staff to conduct evaluation.

As referred to in the second paper, evaluation capacity building (ECB) encompasses the ability to do and use quality evaluation at individual and organisational levels (Cousins et al., 2014; Stockdill et al., 2002). Evaluation knowledge, skills and attitudes are central to ECB (Preskill & Boyle, 2008). Some of these are also key factors requested by the informants: knowledge and understanding about evaluation in general, sense of ownership to the guidelines, communication and transparency about evaluation but also organisational capacity relating to leadership, support structures and evaluation recourses (Cousins et al., 2014). Many of these aspects were perceived as low or almost absent by the academics in this study, which means the evaluation capacity can be considered low. However, the academics called for more time to do evaluation, better support structures, more involvement and feedback, as well as better communication about evaluation, all of which are components of ECB (Borch, 2020). The academics had good intentions for evaluation practice, and they ensured evaluation was carried out within the available time and prerequisites, but they were not provided support structures or a framework that gave them optimal opportunities to carry out what in ECB is labelled as quality evaluations. Another example of low ECB is described in Borch et al. (2021) where descriptive statistics and mean scores of student satisfaction are presented in a report disregarding a response rate of only 20%. This illustrates that there is a need for more knowledge about how to interpret and analyse students' responses. Establishing evaluation capacity and ensuring evaluation practice from which educational quality, teaching and student learning will benefit is hard work and considered to take time. However, in order to strengthen the learning potential of student evaluations, I believe it is worth the time and effort it may take.

The third paper analyses organisational dimensions with the evaluation system: the reporting and documentation of evaluation. We found that the documentation process is not according to intentions described in QAS as the "Quality circle" or interaction circle of planning, action, analysis and improvement. The image of student evaluation practice projected in the reports differed from what the informants described. First, most of information about challenges and poor evaluation results academics described in programme reports and in the interviews were not mentioned in the reports on department and faculty levels. I would argue that it is unfortunate that this information does not reach higher institutional levels. When the leadership and administration get no signals about needs for improvement or examples of good practices others can learn from, it limits the opportunities for strengthening evaluation practice. Second, meaningful and valuable evaluation practice and use described by academics and students in the interviews, like the formative aspects in an ongoing dialogue about students' learning, are

not given attention in the educational quality reports. Nonetheless, these dialogues must not be underestimated. This kind of evaluation seems to influence students' learning processes and their satisfaction the most. With support in the findings of this study we can anticipate that evaluation indeed influences teachers' and students' reflections upon learning and teaching, for a long time after the evaluation is conducted, as suggested by Kirkhart (2000). However, the information about evaluation that academics are asked to report back to the higher institutional level does not concern this kind of use; therefore, readers of these documents may miss out on aspects that actually affect educational quality. When the reports above the programme level (department and faculty) concluded that evaluation practice was carried out according to QAS and internal procedures, this can in itself rather underline a *symbolic use* of evaluation simply to legitimise that student evaluation exists. This is otherwise identified as a ritual (Anderson, 2006) and a requirement in which academics are not engaged (Johansen, 2020). This is contrary to the intention described in QAS, which states that documentation of evaluation and educational quality aims to describe educational quality, identify flaws in quality work and poor educational quality and serve as the basis for improvement measures (UiT, 2012).

Evaluation reports are only one of several mechanisms that may facilitate evaluation use (Patton, 2008). Nevertheless, Harris (2017) argues that such reports should include information that can benefit the stakeholders for which the evaluation was intended for.

As one of the purposes of student evaluation is student empowerment and evaluation is expected to be part of students' learning processes, it is unfortunate that the student perspective in general is missing from the educational quality documents but also in the development of evaluation methods, analysis and implementation of findings. If student empowerment is understood as the *opportunity* to give feedback about their education and *representation* in different boards, one might argue that existing evaluation practices are examples of student empowerment. However, I will argue that student empowerment should involve incorporation of their perspective throughout the evaluation process. Although the European Standard and Guidelines for Quality Assurance (ESG) aimed to increase the role of students as stakeholders in internal quality assurance, little is known about students' real empowerment and roles in these processes (Logermann & Leišytė, 2015). Research on evaluation has shown that students doubt that their opinions matter (El Hassan, 2009; Kite et al., 2015; Spencer & Schmelkin, 2002) and that the student voice is not being heard as intended (Blair & Valdez Noel, 2014). Nonetheless, few scholars seem to provide approaches that actively involve students in

educational evaluation and internal quality assurance (Giles, 2004; Stalmeijer et al., 2016). Research within the field of curriculum design has, however, suggested several benefits with and approaches to involving students in pedagogical planning (Bovill et al., 2011; Brooman et al., 2015), which may also be a recommendation for planning of evaluation. The study by Bovill, Cook-Sather and Felten (2011, p. 137) involved students as co-creators of curricula, including evaluation, and concluded that the involvement "helped students to realise that they were being taken seriously and that their participation was meaningful rather than tokenistic".

The students at UiT are, as mentioned, not involved in the development of evaluation questions, analysis of the findings, dialogue about the results or in the reporting of evaluation use in educational quality reports. However, students at UiT called for feedback about evaluation results and whether these have been used (Borch et al., 2020; Hoel & Dahl, 2018). If we succeed at increasing active student involvement in evaluation, the sense of ownership in evaluation and opportunity to provide meaningful feedback may increase. When the local QAS was approved by NOKUT, the committee suggested involving students more actively in quality assurance to strengthen their empowerment (NOKUT, 2012). Based on the findings in the present study it seems like actual student empowerment is still low. Educational evaluation in general, and in this study in particular, does not seem to be inspired by fourth- generation evaluation approaches wherein stakeholder involvement throughout the evaluation processes is regarded as a key element for optimal evaluation use, meaningful approaches and a sense of ownership over evaluation. If we manage to establish dialogues about how students perceive their learning processes and also take their feedback into consideration, this can be regarded as a step towards more active student empowerment showing students that their opinions matter.

6.2 Pedagogical dimensions with student evaluation practice

Based upon the results in the study, I will in this sub section of the discussion focus on pedagogical dimensions with student evaluation and the student learning perspective in evaluation. I would like to return the reader's attention to the UiT's QAS that states: "evaluation is part of students' learning processes" (UiT, 2012). As QAS does not provide any further information about *how* student evaluation can be incorporated as part of student's learning processes, it is open to interpretation. From my point of view should evaluations that are part of students' learning processes as a minimum invite students to provide feedback about how the teaching and learning activities contributed to their learning processes.

Scholars have proposed that we should stop regarding students as consumers of education and student satisfaction as a measure of educational quality if we would like to understand evaluation as an approach that can increase student learning and not just promote accountability (Leckey & Neill, 2001; Schuck et al., 2008). By measuring satisfaction and not dimensions with courses that promote and hinder learning, and viewing students as consumers not active learners, increased educational quality is not likely obtained through evaluation. Nonetheless, it is possible to change the focus of evaluation processes and embed more formative evaluation activities that ask students about how the courses and teaching to help students develop insight and knowledge are aligned with the intended learning outcomes (Andersson et al., 2012). The findings in this study indicated that there *are* evaluation practices at all the included programmes but that not all courses that facilitate students' reflecting upon their learning processes, yet the learning from these evaluation approaches was not described in the reports. As these evaluation approaches contributed more to adjustments of teaching than surveys (Borch et al., 2020) this evaluation approach can be strengthened at the university in order to improve educational quality. Student evaluation practice has a great potential for developing reflective dialogues about learning processes and can be approached in different ways (Roxå et al., 2021; Ryan, 2015).

Academic developers (AD) are involved in a wide range of activities, wherein providing support to academics in student evaluation practice is just one of many (Amundsen & Wilson, 2012; Gibbs, 2013). ADs are expected to serve academics and university management and promote educational quality through their activities. ADs' "mediating" position in the organisation has given them roles as change agents (Cordiner, 2014; McGrath, 2020). Active support from ADs is recommended in evaluation processes to shift the focus in evaluation from teaching to student learning (Gibbs, 2013) and suggested as a factor in improving student learning as a result of evaluation (Darwin, 2012). The academics in this study expressed a need for support in evaluation processes. They requested more attention about student evaluation in general and said they wished for a more shared practice throughout the evaluation process. They draw a picture of evaluation as a lonely and private part of their job because they never discuss evaluation with their colleagues or get feedback about the data they collected from their leaders (Borch, 2020). The academics in this study did not specify who they wanted support from but based upon previous research I suggest that there is potential at the university for ADs to take a more active role in evaluation practice in order to facilitate educational quality, as change or

quality agents. Andersson, Ahlberg and Roxå (2012) found that while the faculty and programme leadership regarded summative evaluation as useful for quality assurance and quality enhancement, the teachers benefited more from formative evaluation because the immediate feedback from the students could help them adjust the teaching in order to increase students' learning. One may say that evaluation is only formative if the teachers review and reflect upon the results in ways that will affect their teaching and professional development (Hobson & Talbot, 2001). Tavares et. al. (2017) found that academics' perceived participation in internal quality assurance increased their awareness of educational quality without leading to improved teaching. From my perspective, it is not enough to conduct student feedback during courses without following up the results. Especially if students provide feedback that indicates that there were elements of the teaching that hindered their learning processes, students in particular expect teachers to take an active stance on how they can use students' responses (Kember et al., 2002).

ADs can support teachers and encourage them to embed formative student evaluation in their courses, either as a topic in pedagogical courses or in close interaction with the teachers and/or course leader. Literature reviews also find that teachers' understanding and active use of the evaluation results increase when they get support in interpreting and analysing the data (Hampton & Reiser, 2004; Penny & Coe, 2004; Piccinin et al., 1999; Piccinin & Moore, 2002). This can be considered a collaborative approach to evaluation. I would like to underscore that I do not regard any increased use as an aim in itself, rather increased use for intended purposes. Studies suggest that consultation to assist and support teachers provided by academic developers will increase the use of the data in teaching improvement (Kogan et al., 2010; Lang & Kersting, 2007; Nasser & Fresko, 2002; Vasey & Carroll, 2016).

I would like to underline that it is neither the established regulation of student evaluation nor the format in itself that is decisive for whether evaluation can be integrated into parts of students' learning processes but rather aspects with how evaluations are carried out and followed up.

Harvey (2003, p. 4) emphasises that it is more important that the universities' quality assurance practice have "action cycles" that ensure evaluation is used to enhance student learning rather than having evaluation systems to collect data in place. By "action cycles" he refers to continuous circles consisting of analysis, reporting, action and feedback (ibid). The findings in

this study show that the quality assurance system contains guidelines for evaluation practice that, if followed, could ensure feedback to students and teachers about the results (Borch, 2020). However, the feedback about evaluation results to the students seems to be scarce at the included programmes. This was also the case when NOKUT approved the local QAS: they recommended strengthening the feedback loops within the university (NOKUT, 2012). Establishing mechanisms for feedback to students and teachers may contribute to increasing the dialogue and reflection about evaluation, which in turn can enhance use of evaluation for educational development (Harvey, 2003; Shah et al., 2017; Watson, 2003). In order to use student evaluation to improve teaching, and hence student learning, Harvey (2003) suggests that teachers spend more time with the students and discuss how they can provide thoughtful feedback to their teachers, which in turn will develop metacognitive skills. Evaluative dialogues and collaboration between students and academics in evaluation processes can increase reflection upon learning and facilitate shared ownership (Freeman & Dobbins, 2013).

ADs can help establish collaborative approaches between teachers and students in pedagogical planning, curriculum design, including evaluation and reflection on learning. The authors and ADs Bovill, Cook-Sather and Felten (2011, p. 143) suggest that involving students in such collaboration "is a significant step in deepening engaged learning and might, therefore be understood as professional responsibility for academic developers". Involving students as partners or "legitimate actors" in academic development is not a very common practice, yet a desirable collaboration (Felten et al., 2019, p. 195). Stein et al. (2020) suggest involving students and teachers in a partnership when developing student evaluation practice for the purpose of increasing students' engagement with and learning focus in evaluation.

In order to overcome the perception of evaluation as a private practice and the uncertainty about responsibilities the academics described, I suggest striving for establishing a sense of ownership over evaluation among the academics. Possible approaches to establishing shared understandings of evaluation might be found in literature about collegial reflections on educational quality (Ramsden, 2003; Schuck et al., 2008). We know from a review study on evaluation use that a key principle to ensuring use of evaluation is to engage stakeholders or users of evaluation so that they establish a sense of ownership over the evaluation (Johnson et al., 2009). The correlation between evaluation and improved student learning is highly dependent on the active intervention of academic development or supplementary evaluative

strategies (Darwin 2012). As stated in the introduction of the thesis, student evaluation should always be balanced with other indicators and approaches to promote educational quality.

6.3 Organisational and pedagogical dimensions in student evaluation practice: system alignment

Edström (2008) proposed the concept of *system alignment* as support for analysing different processes at a university from a student-learning perspective. The concept is presented in the literature review section 2.5. In the following, I am using system alignment as a starting point of reference to illuminate to what extent and how the student learning perspective is embedded in evaluation practice at the university by applying the concept to my empirical findings. To start with, I will describe how I consider the evaluation system to have a teacher- and teaching-focus rather than a student and student-learning focus today. Next, I will present a framework for evaluation practice where the student-learning perspective is embedded at the institution throughout the whole process. This framework is my interpretation, development and visualisation of the system alignment concept proposed by Edström. I would like to emphasise that I do not regard the framework as a standard or a template, but rather a *guiding* framework for enhancing the learning perspectives in evaluation. By proposing the framework based upon my findings, I aspire to visualise how organisational and pedagogical dimensions and different actors interact with each other when student evaluation practice is carried out.

6.3.1 Low student learning perspective in today's evaluation practice

It seems to be a misalignment between intention of student evaluation and the actual student learning focus in evaluation practice at UiT. This misalignment and low student learning focus were found in 1) the written evaluation methods, 2) the translation of evaluation, and 3) the documentation of student evaluation practice and educational quality.

Although scholars back in the mid-1990s claimed that higher education had been through a transformation from being teaching- and teacher-focused to learning- and student-centred (Barr & Tagg, 1995), this transformation seems to be delayed when it comes to evaluation. Ramsden (2003, p. 220) underlines that evaluation needs to be better aligned with other teaching and learning activities: "…we must evaluate in a way that coheres with the principles of good teaching, learning and assessment", and used the terms "coherent" or "aligned" evaluation.

The proposed framework for how today's evaluation practice may improve (Figure 3) is based upon my findings and is also in coherence with the existing national and European regulation of quality assurance. The national regulation concerning quality assurance and quality development in higher education section 1-4 (Studiekvalitetsforskriften, 2018) states that institutions must use knowledge obtained through quality assurance to enhance the quality of the institutions' study programmes and uncover instances of deficient quality. The European guidelines—The Standards and Guidelines for Quality Assurance in the European Higher Education Area (ESG) set by The European Association for Quality Assurance in Higher Education (ENQA)—encourage institutions to take a student-centred and learning-outcome-focused approach to teaching and internal quality management (Introduction by Walsch, the ENQA president in Bergsmann et al., 2016, p. 7; ENQA, 2015).

6.3.2 Systemically aligned evaluation practice

The framework I present in this section was inspired by system alignment thinking and my empirical data to illustrate how different pedagogical and organisational dimensions interact with each other in evaluation practice. It is important to underline that I do not regard the framework as a simplistic process or as an exhaustive model that captures all components and actors that influence evaluation processes; moreover, that the different evaluation stages depicted in the framework are overlapping each other. I would like to emphasise that interaction with other actors and different aspects, e.g., backgrounds, traditions, relations, contexts and policies affect the actors' translation of evaluation processes. This is illustrated and discussed in paper two (Borch, 2020).

I have developed two figures applying the concept of system alignment to my empirical data. Figure 2 summarises my findings concerning the actual practice of student evaluation today. Figure 3 depicts an evaluation system in which student learning plays a greater role than it currently does.

In both figures human actors, particularly students, academics and administrative staff, play important roles in constructing evaluation practice throughout the evaluation process. Figures 2 and 3 include symbols representing the different stakeholder groups. The presence or absence of these symbols in different fields of the figures is significant. For instance, the symbol for

students is absent from the field labelled *evaluation reporting and feedback* in Figure 2. This indicates that students in this study are not directly involved in this stage of the evaluation process. The size of the symbols is also of importance as it represents the extent to which the corresponding stakeholder groups contribute in the different evaluation stages. The upper part of the figures shows arrows that illustrate a process of continuous analysis, feedback and improvement of teaching and learning that *intentionally* in QAS shall frame evaluation practice at the university (UiT, 2012). Inside this circle are different versions of evaluation practice taking place, some illuminated by the informants in this study, but also other practices, all affected by cultures, other actors, values, traditions and policies in the contexts in which they are carried out in. By taking a Utilization Focused Evaluation approach to evaluation, the intended purpose(s) shall be embedded throughout the evaluation processes, from the planning of an evaluation, in regulation and up to the documentation and feedback to the students and teachers because evaluation use depends on what happens in the different stages of an evaluation process (Patton, 2008, 2011). I aimed to illustrate a systemic approach to evaluation as processes wherein the student- and student-learning perspective is central in Figure 3.

6.3.3 Today's evaluation practice

Below I am presenting a framework illustrating today's evaluation practice (Figure 2)[27] based upon the findings in this study.
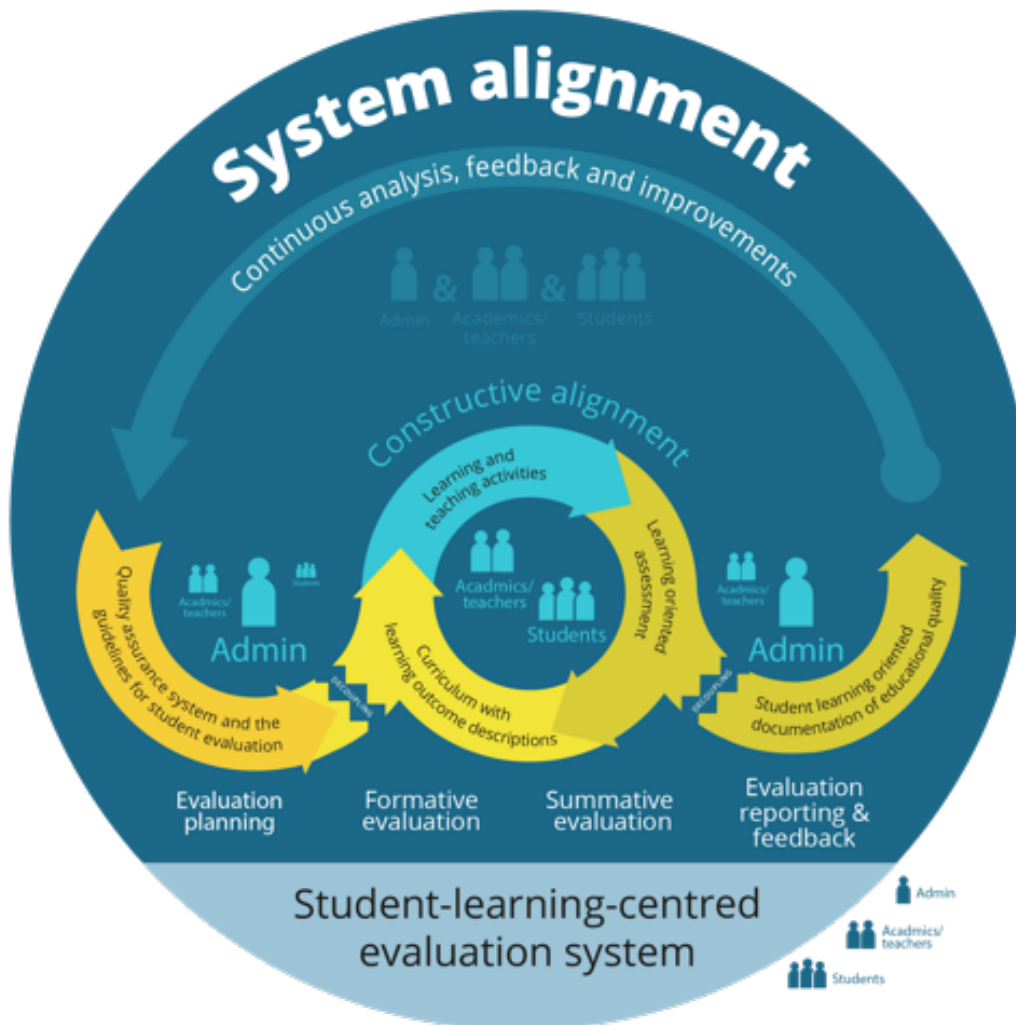

Figure 2 Today's student evaluation practice

Figure 2 illustrates that the administrative contribution dominates in the development of the quality assurance system—teachers are expected to follow. From the 1980s to 2015, the number of administrative staff increased more than academics in Norwegian higher education (Gornitzka & Larsen, 2004; NOU, 2015). In the same period, reforms based on principles in NPM and managerialism contributed to more accountability and control of higher education (Bleiklie & Frølich, 2014). Administrative staff at UiT were central in developing and following

---

[27] A bigger sized framework is found in Appendix 7

up on QAS, while the academics' contribution to the development of the evaluation guidelines in the quality assurance system was low. Despite an increased focus on accountability and control in quality assurance processes, the local QAS states that student evaluation is part of students' learning processes. My research found that in the translation of the evaluation guidelines into practice, a disconnection or decoupling, visible in Figure 2, appears. This is elaborated on in paper two (Borch, 2020). In order for evaluation to be student learning-centred and embedded in students' learning processes, organisational dimensions and communication about evaluation internally at the university should also encompass this perspective, e.g., by involving students more actively in evaluation processes and establishing a more collaborative approach to evaluation among the involved stakeholders. Today's practice of evaluation diverges from the intentions in many ways, particularly because students are not actively involved in the development of evaluation questions, methods, analysis of results or in implementation of findings and because academics are left to themselves throughout the evaluation process. The informants described that there is no collective engagement about evaluation at programme level or across the organisational levels. QAS is simply communicated on the webpage but no further information or courses about evaluation, recommendations, expectations, etc. is provided to the academics. Nonetheless, some of the departments have established routines wherein administrative staff send academics reminders to conduct student evaluation and a request to provide evaluation results for educational quality reports. This in itself may strengthen the control dimension with evaluation.

When academics are left to themselves in evaluation processes without any counselling or support and do not have sufficient time to engage in evaluation, symbolic evaluation may easily be carried out but probably not evaluation suitable for educational development. In the present study, this seems to be unintentional. Nevertheless, dialogue-based evaluation practice where students and academics are the main actors exists. Examples of such evaluation practices are found in the middle circles of Figures 2 and 3. These practices seem to be more used in educational development, but this evaluation approach is not the dominant one. I would like to underscore that a diversity of student evaluation practices and different types of uses are carried out and found in today's evaluation practice. Johansen (2020) found in a study from UiT that academic programme leaders at three faculties other than the one in this study perceived quality assurance, including evaluations, as pro forma activities controlled by administrative staff they themselves had low sense of ownership of. The present study also found that administrative staff played central roles in how evaluation was carried out and reported, whereas the academics

are not actively involved throughout evaluation processes, particularly in development of the formal evaluation system. This may affect the sense of ownership over evaluation practice for the academics and contribute to an uncertainty about whose responsibility it is to follow up on evaluations. It may therefore not be a surprise that evaluation results are not used more actively in educational development.

Research has shown that systematic implementation of QA can benefit student ratings positively (Barrie et al., 2005), a more systemic approach to student evaluation and a strategy for how to achieve the intended outcome based on theory of change is suggested by scholars in the field (Roxå et al., 2021). The findings in this study, however, showed that the academic leaders requested better organisational systems and structures to follow up on evaluations (Borch, 2020). No measures exist for how to improve evaluation practice in evaluation documents (Borch et al., 2021) and the intended objectives of evaluation were not achieved (Borch et al., 2020). Moreover, Figure 2 illustrates that evaluation practice is not part of a continuous process of analysis, feedback and improvement. As described in paper three (Borch et al., 2021) the student perspective was absent in the reports on higher institutional levels (department and faculty) and there was limited or no information about weaknesses and strengths with the programmes. The reports merely reported or confirmed that the programmes had a system in place, rather than providing descriptions of challenges with today's evaluation practice and the educational quality perceived by students and academics. If institutions shall be able to improve evaluation practice and carry out measures that can support improvement of educational quality these reports contain information that guides improvement. If not, such information must be communicated to higher levels in other ways. Today useful information "disappears" between programme and department level and can therefore hinder organisational learning. This is illustrated as decoupling between summative evaluation and evaluation reporting in Figure 2. Within translation theory different modes of translation are identified: when an idea is translated in ways that differ from the original idea, one might say that an idea gets decoupled in the translation process (Røvik, 2014). This was described in the second paper as modified translation, characterised as unintentional decoupling.

There may be different explanations as to why this decoupling happens. One may relate to the first decoupling, described in paper two and shown in the figure, between evaluation planning and formative evaluation. The second decoupling happens at different levels, by the academics at programme level and by administrative staff at department and faculty level. On the

programme level, reports about educational quality including descriptions of evaluation practice are written for only half of the programmes. At the remaining programmes, leaders said they either shared no information about evaluation practice or the administrative staff had direct access to evaluation results in the feedback platform Questback. The administrative staff on the department level probably had limited written information about student evaluation from the academics on which to base the departmental reports. When some programmes do not pass information to the next level in the first place because some of the leaders do not write reports with analysis of evaluation findings and suggestions for action, this probably has different explanations. It may relate to limited knowledge about the system in general—they simply do not know that written documentation is required. Another possible explanation that is not directly expressed by the academics in this study—but by other scholars—is that they do not follow up on these routines because they do not regard quality assurance as an important part of their job (Ese, 2019), particularly "feeding" the system with reports (Anderson, 2006). Two of the informants in this study gave some explanations to why the system is not followed closely that may relate to the findings of Ese (2019) and Anderson (2006). One said that student evaluation unfortunately was not a priority in busy times. Another academic said that he had colleagues who had stopped evaluating and creating reports because nobody had ever shown interest or requested these reports (Borch, 2020). This underscores the importance of establishing feedback loops within the university and between different actors in evaluation processes. The diverging picture of evaluation practice described by the informants and in the educational quality reports on programme level from the one presented in the reports on department and faculty level must be further explored. The reports on department and faculty level are written by administrative staff. As I do not have first-hand knowledge from the administrative staff, it is necessary to explore further why documentation happens the way it does.

Documentation procedures at UiT aim to inform stakeholders about educational quality and evaluation findings. Information from one organisational level is meant to be shared with higher organisational levels, e.g., from programme to department and from department to faculty (UiT, 2012). My study shows that reports are merely descriptions of how quality assurance is carried out rather than how it affects educational quality. This may relate to what kind of information is requested from higher institutional levels. Analysis of these questions show that they were not questioning information about the perceived educational quality, nor strengths or weaknesses with the programmes but several questions about compliance with the system

(Borch et al., 2021). Information about evaluation practice found in the reports cannot be used to improve evaluation practice. If evaluation is regarded as part of student learning processes and student empowerment, it is unfortunate that the student perspective on educational quality is missing from the reports, as well as in development and follow-up on evaluation.

6.3.4 Systemically aligned evaluation practice: An implication for practice?

In the proposed Figure 3, I have developed a framework that might be seen as a possible implication for practice[28]. The aim is to provide a *guiding* framework for institutions who would like to strengthen a student-learning focus in evaluation practice. However, there are different ways in which the framework can be carried out and these need to be further explored in practice. The main emphasis in the framework is an integrated student and learning perspective throughout evaluation processes on different organisational levels.
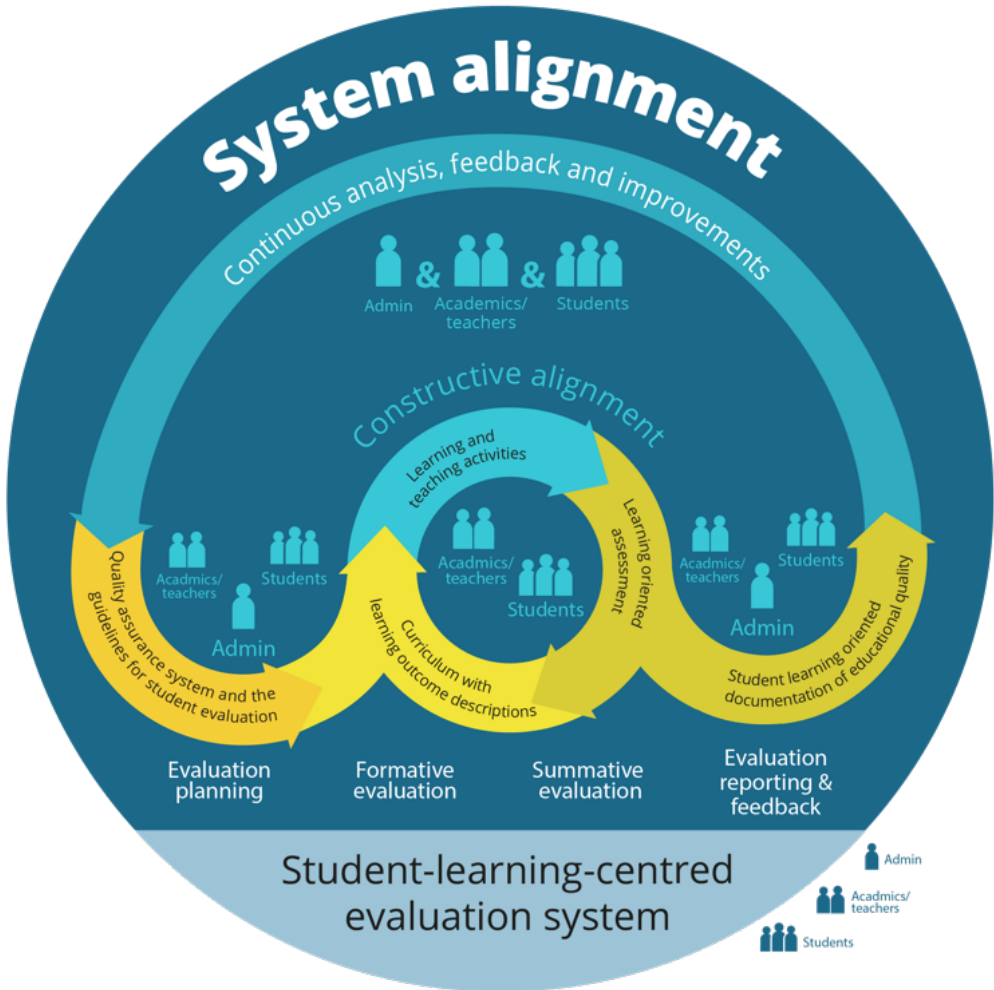

Figure 3 Systemically aligned evaluation practice

---

[28] A bigger sized framework is found in Appendix 8.

In this framework of a systemically aligned evaluation practice, I have included students and teachers throughout the evaluation process as an attempt to strengthen the student and learning perspectives in student evaluation. The administrators are still actors in evaluation processes but are in collaboration with teachers and students. I will not provide approaches about how to accommodate a change in evaluation understanding. My intention in proposing a framework is to shed light on different dimensions that affect how evaluation is carried out. I would like to underscore that it is necessary to actively involve the key stakeholders throughout the whole process to better balance the multiple purposes in student evaluation and strengthen dimensions of evaluation that promote enhancement of educational quality and student learning. The formative dimensions of evaluation can be increased and more attention should be given to what students and academics consider meaningful evaluation practices that increase reflection upon learning processes (Borch et al., 2020). In Figure 3, different stages in the evaluation processes are aligned with each other and there are no disconnections throughout the process. The continuous process of analysis, feedback and improvement is also strengthened in a possible systematically aligned student evaluation process in the future.

I acknowledge the importance of quality assurance and accountability in educational quality reporting, particularly for reports to institutional boards and quality assurance agencies. However, the reports in this study are from programme, department and faculty levels. If it is expected that there should be alignment in information about educational quality across organisational levels and that evaluations contribute to achieving different purposes, all perspectives should also be appearing in the reports. I suggest integrating both students' and teachers' perspectives and reflections on educational quality in future reports. Quality assurance is an important issue for the institution and therefore also in the educational reports; however, for the teachers and students, quality enhancement has a higher priority (Andersson et al., 2012). Student evaluation has existed at the university and in the sector since long before it became mandatory. Although evaluation practice in the beginning was self-driven and voluntary, it still took place, though without formalised systems that regulated evaluation and required documentation.

In order to create systemically aligned evaluation practices, an evaluation capacity (ECB) must be established. This will take time and effort. In evaluation capacity building, the principles are based on research on evaluation use. Essential to accommodating ECB, is increased stakeholder

involvement, transparency about evaluation, clarity about whose responsibility it is to plan, carry out, analyse and follow up on evaluation and, moreover, increased evaluator competence. I believe that strengthening students' and academics' involvement and perspectives throughout the evaluation process as illustrated in Figure 3 above is a start also to establishing ECB. Some ways to involve stakeholders can be found in fourth-generation evaluation approaches like Participatory, Collaborative and Utilization-focused evaluation wherein the key stakeholders are involved throughout the evaluation processes (Fitzpatrick, 2012). For example, in discussions about the aim with evaluation to make sure there is a shared agreement and understanding about the evaluation purpose, in the development of evaluation questions in order to create meaningful questions, in analysis of the findings, etc. By increasing the involvement of students and academics throughout evaluation processes the university management can show central stakeholders that they emphasise evaluation as part of students' learning processes. I would like to underscore that I do not regard it as sufficient to develop a system or a policy and expect these to be followed. Saunders, Trowler and Bamber (2011, p. 208), the editors of the book "Reconceptualising Evaluation in Higher Education – The Practice Turn", state that "there is a clear distinction between policy-in-text and the policy enactment process". The gap between formal organisational systems and actual practice has been noticed by scholars within New Institutionalism for decades (Brunsson, 2003; Meyer & Rowan, 1977). After evaluation became mandated in many public sectors and the number of internal self-evaluations grew, critics pointed to how evaluation has become "popularised" with increased stakeholder involvement and a growing "expectation that lay persons will become good evaluators based on a short introduction and rudimentary training in methodology" (Dahler-Larsen, 2006, p. 145). Nonetheless, in the proposed framework, my recommendation is that individuals should not be left to themselves in evaluation without much knowledge about evaluation. Evaluation practice should rather be carried out in collaboration with different stakeholders, guided by the intended purpose of evaluation and informed by research on student evaluation and evaluation use. As long as today's practice continues unchanged, the pedagogical dimensions of student evaluation practice will most likely continue to be backgrounded for accountability and control.

6.4 Limitations of project and further research needs

I have in this study explored student evaluation practice from the perspectives of students and academics and provided my interpretation of documents from a textual agency perspective. However, as I acknowledge that actors construct their social realities, including student

evaluation practice, the findings would have been different if the topic had been explored with other actors in other contexts. The project could have been strengthened with first-hand knowledge from the administrative perspective, e.g., by interviewing administrative staff and university management. In this project, documents written by administrative staff are analysed by me as a researcher; thus, I have included work that indirectly says something about the administrative role and involvement in evaluation processes.

The literature on quality work lacks knowledge about how different actors interact with each other in student evaluation practice and how pedagogical and organisational dimensions may affect how evaluation can be carried out to promote learning. My study contributes to the field; however, as the study is a small-scale and explores student evaluation practice at eight health profession education programmes from the same faculty and institution, there is a need for more studies in the future, as well as from other faculties and institutions and observation studies. The proposed framework for a more student-learning-centred evaluation system needs to be further explored because there are different approaches to how the framework can be carried out in practice.

7.0 Concluding remarks

This study has explored how student evaluation is carried out and used by central actors at UiT. By drawing upon different empirical data and analytical perspectives, it became possible to gain insight about pedagogical and organisational dimensions that affect how evaluation practice was carried out at different organisational levels. My analysis found dimensions within the organisation that seemed to limit use of evaluation for pedagogical purposes. Some of these dimensions relate to how the evaluation system and guidelines are developed, communicated and followed up by different actors and how the actors interact across organisational levels. Figure 2 summarises how student evaluation practice at the university today is carried out. I have named the visualisation of evaluation practice in the figure(s) as a "systemically aligned evaluation" framework. The framework is based upon an understanding of student evaluation as part of constructive alignment and the concept "system alignment" proposed by Edström (2008). The framework visualises how student evaluations are processes involving interaction between different stakeholders across organisational levels. Some actors, particularly administrative staff, are in today's evaluation practice (figure 2) given more central roles than others in the development of the system and reporting of evaluation practice. Students and

academics are key actors in the formative and summative evaluation. However, in the reporting of this practice at department and faculty level, the student perspective is absent and the academics' perspective is not given much attention. The findings in this study illustrate that evaluation is understood differently by stakeholders at different levels. Saunders (2011, p. 205) wrote that there are "different ways of seeing" educational quality and the approaches that ensure and enhance the quality of education. Students and academics expect that student evaluation results would contribute to educational improvement. However, this study shows that evaluation practice has the potential to improve, particularly in embedding evaluation as part of student learning processes and strengthening formative evaluations, which in turn may increase use of quality improvement. In this study, the dialogue-based evaluations seem to have a stronger learning focus than the written evaluations and can be seen as a viable alternative to written evaluations. If the student learning perspective shall be incorporated throughout the evaluation process, I suggest that the interaction between academics, students and administrative staff should be better balanced. Further, collaborative dialogues between the involved stakeholders may contribute to a shared awareness of the purpose of student evaluation. Figure 3 can be seen as a *guiding* framework and a possible implication for practice for universities that aim to embed a student learning perspective in evaluation. This may help them develop evaluation practices that contribute to educational quality enhancement.

There seems to be a gap between the intended use and practice articulated in QAS and use described by academics and students. I identified aspects of the evaluation practice that limit use of evaluation for quality enhancement purposes. Some of these limiting aspects presented in the papers and in the discussion section of the thesis include: judgmental, teaching-focused evaluation questions, low ECB and the lack of established evaluation plans that included involvement of students and academics. In short, the university seems to have developed a system that many of the involved stakeholders do not have capacity to follow up. My study does not find a continuous process of feedback, analysis and implementation of evaluations, nor existing implementation plans. The approaches to evaluation are highly dependent on individual translations rather than details in internal evaluation guidelines or the quality assurance system.

Students are qualified to give feedback about their own learning experience, their motivation, the perceived difficulty and achievement, how the teachers interact with students, to what extent teachers' feedback is valuable and help them in their learning processes, their engagement etc.

They are, however, not experts in course curriculum design nor relevance of course content for professional development, and they might know nothing about course adjustments the teachers made based upon previous student feedback, exam results, etc. I argue that there is a need for a balanced evaluation system with different methods, because every single evaluation method has its limitations. In addition to student evaluation, I suggest there is a need for quality assurance systems that base their judgment and enhancement of educational quality on multiple sources like teachers' self-evaluation, peer observation, evaluation of course material by colleagues, formal and informal feedback as well as informal measures and maintenance described in quality work literature. Each source is a topic complex enough for a PhD project in itself.

I would like to return to the example presented in the introduction of the thesis and my first meeting with student evaluation. Hopefully the findings in this project will contribute to further development of evaluation practice in ways that make evaluation meaningful and provide insight about teaching, courses and programmes that can be useful in educational development. If I knew what I now know about evaluation when I created my first programme evaluation eight years ago, the evaluation approach I took would have been different.

As student evaluation is mandatory in Norway, it is not something we can ignore but the institutions can choose how to approach it. By acknowledging that internal evaluation has different functions and purposes than external evaluations, academic leaders and teachers can establish evaluation practices that are well suited to gaining insight about students' learning processes and exploring how the teaching approaches contributed to or inhibited learning. By approaching internal student evaluation as learning-centred at all organisational levels and facilitating interaction between the involved stakeholders throughout evaluation processes, evaluation may be re-established as pedagogical practice. If used consciously, student evaluation has great potential to contribute to improved educational quality and lead to increased student learning.

References

Aarstad, J. (2012). Studentevalueringer i høyere utdanning: hva kan den internasjonale forskningslitteraturen lære oss? *Uniped*(01), 34–45.

Abrami, P. C. (2001). Improving judgments about teaching effectiveness using teacher rating forms. *New Directions for Institutional Research*, *2001*(109), 59–87. https://doi.org/10.1002/ir.4

Abrami, P. C., d'Apollonia, S., Rosenfield, S., Perry, R. P., & Smart, J. C. (2007). The dimensionality of student ratings of instruction: What we know and what we do not. In: Perry, R. P., Smart, J.C. (Eds.), *The Scholarship of Teaching and Learning in higher education: An evidence-based perspective.* Springer. https://doi.org/10.1007/1-4020-5742-3_10

Ackerman, D., Gross, B. L., & Vigneron, F. (2009). Peer observation reports and student evaluations of teaching: Who are the experts? *Alberta Journal of Educational Research*, *55*(1), 18–39.

Adams, M., & Umbach, P. (2012). Nonresponse and online student evaluations of teaching: Understanding the influence of salience, fatigue, and academic environments. *Journal of the Association for Institutional Research*, *53*(5), 576–591. https://doi.org/10.1007/s11162-011-9240-5

Afonso, N. M., Cardozo, L. J., Mascarenhas, O., Aranha, A., & Shah, C. (2005). Are anonymous evaluations a better assessment of faculty teaching performance? A comparative analysis of open and anonymous evaluation processes. *Family Medicine*, *37*(1), 43–47.

Ahmadi, M., Helms, M. M., & Raiszadeh, F. (2001). Business students' perceptions of faculty evaluations. *International Journal of Educational Management*, *15*(1), 12–22. https://doi.org/10.1108/09513540110366097

Albaek, E. (1995). Between knowledge and power: Utilization of social science in public policy making. *Policy sciences*, *28*(1), 79–100.

Alderman, L., Towers, S., & Bannah, S. (2012). Student feedback systems in higher education: A focused literature review and environmental scan. *Quality in Higher Education*, *18*(3), 261–280. https://doi.org/10.1080/13538322.2012.730714

Alkin, M. C., & King, J. A. (2016). The historical development of evaluation use. *American Journal of Evaluation*, *37*(4), 568–579.

Alkin, M. C., & King, J. A. (2017). Definitions of evaluation use and misuse, evaluation influence, and factors affecting use. *American Journal of Evaluation*, *38*(3), 434–450. https://doi.org/10.1177/1098214017717015

Alvesson, M. (2003). Methodology for close up studies – Struggling with closeness and closure. *The International Journal of Higher Education and Educational Planning*, *46*(2), 167–193. https://doi.org/10.1023/A:1024716513774

Alvesson, M., & Sköldberg, K. (2017). *Tolkning och reflektion: Vetenskapsfilosofi och kvalitativ metod* (3rd ed.). Studentlitteratur.

Alvesson, M., & Sköldberg, K. (2018). *Reflexive methodology: New vistas for qualitative research* (3rd ed.). SAGE Publications.

Amundsen, C., & Wilson, M. (2012). Are we asking the right questions? A conceptual review of the educational development literature in higher education. *Review of Educational Research*, *82*(1), 90–126.

Andersen, H. L. (2010). "Constructive Alignment" og risikoen for en forsimplende universitetspædagogik. *Dansk Universitetspædagogisk Tidsskrift*, *5*(9), 30–35.

Andersen, H. L., & Søndergaard, L. (2006). Undervisningsevaluering som redskab til kvalitetsudvikling af undervisningen. *Dansk Universitetspædagogisk Tidsskrift*, *1*(2), 22–27.

Anderson, G. (2006). Assuring quality/Resisting quality assurance: Academics' responses to 'Quality' in some Australian universities. *Quality in Higher Education*, *12*(2), 161–173. https://doi.org/10.1080/13538320600916767

Anderson, G. (2008). Mapping academic resistance in the managerial university. *Organization*, *15*(2), 251–270.

Andersson, R., Ahlberg, A., & Roxå, T. (2012). Summative quality assurance systems: Not good enough for quality enhancement. In C. S. Nair (Ed.), *Enhancing Learning and Teaching through Student Feedback in Engineering* (pp. 93–105). Elsevier.

Arthur, L. (2020). Evaluating student satisfaction - Restricting lecturer professionalism: Outcomes of using the UK National Student Survey Questionnaire for internal student evaluation of teaching. *Assessment & Evaluation in Higher Education*, *45*(3), 331–344. https://doi.org/10.1080/02602938.2019.1640863

Ashwin, P., & Boud, D. (2015). *Reflective teaching in higher education*. Bloombury Academic.

Atkinson, P., & Coffey, A. J. (2011). Analysing documentary realities. In *Qualitative Research* (3rd ed., pp. 77–92). SAGE Publications.

Avery, R. J., Bryant, W. K., Mathios, A., Kang, H., & Bell, D. (2006). Electronic course evaluations: Does an online delivery system influence student evaluations? *The Journal of Economic Education*, *37*(1), 21–37.

Bacon, D. R., Johnson, C. J., & Stewart, K. A. (2016). Nonresponse bias in student evaluations of teaching. *Marketing Education Review*, *26*(2), 93–104. https://doi.org/10.1080/10528008.2016.1166442

Balam, E. M., & Shannon, D. M. (2010). Student ratings of college teaching: A comparison of faculty and their students. *Assessment & Evaluation in Higher Education*, *35*(2), 209–221.

Ball, C. (1985). What the hell is quality? In C. Ball & D. Urwin (Eds.), *Fitness for Purpose: Essays in Higher Education* (pp. 96–102). Society for Research into Higher Education & NFER Nelson.

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.

Bamber, V. (2011). Evaluative practices and outcomes: Issues at the institutional level. In M. Saunders, P. Trowler, & V. Bamber (Eds.), *Reconceptualising evaluation in higher education: The practice turn* (pp. 165–170). McGraw-Hill Education.

Bamber, V., & Anderson, S. (2012). Evaluating learning and teaching: Institutional needs and individual practices. *International Journal for Academic Development*, *17*(1), 5–18. https://doi.org/10.1080/1360144X.2011.586459

Bamber, V., & Stefani, L. (2016). Taking up the challenge of evidencing value in educational development: From theory to practice. *International Journal for Academic Development*, *21*(3), 242–254. https://doi.org/10.1080/1360144X.2015.1100112

Barbour, R. S. (2001). Checklists for improving rigour in qualitative research: A case of the tail wagging the dog? *BMJ*, *322*(7294), 1115–1117.

Barbour, R. S. (2005). Making sense of focus groups. *Medical Education*, *39*(7), 742-750. https://doi.org/10.1111/j.1365-2929.2005.02200.x

Barbour, R. S. (2008). *Doing Focus Groups*. SAGE Publications.

Barr, R. B., & Tagg, J. (1995). From teaching to learning—a new paradigm for undergraduate education. *Change: The Magazine of Higher Learning*, *27*(6), 12–26.

Barrie, S., & Ginns, P. (2007). The linking of national teaching performance indicators to improvements in teaching and learning in classrooms. *Quality in Higher Education*, *13*(3), 275–286.

Barrie, S., Ginns, P., & Prosser, M. (2005). Early impact and outcomes of an institutionally aligned, student focused learning perspective on teaching quality assurance. *Assessment & Evaluation in Higher Education*, *30*(6), 641–656.

Basow, S., Codos, S., & Martin, J. (2013). The effects of professors' race and gender on student evaluations and performance. *College Student Journal*, *47*(2), 352–363.

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

Bennett, L., & Nair, C. S. (2010). A recipe for effective participation rates for web-based surveys. *Assessment & Evaluation in Higher Education*, *35*(4), 357–365. https://doi.org/10.1080/02602930802687752

Benton, S. L., & Young, S. (2018). Best practices in the evaluation of teaching. Idea Paper# 69. *IDEA Center, Inc.*

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.

Berger, P. L., & Luckmann, T. (1966). *The social construction of reality, a treatise in the sociology of knowledge*. Penguin Books.

Berger, P. L., & Luckmann, T. (2000). *Den samfunnsskapte virkelighet*. Fagbokforlaget.

Bergfjord, O. J. (2014). Studentevaluering i høyere utdanning – en empirisk studie fra HiB. *Uniped*, *37*, 33–43. https://doi.org/10.3402/uniped.v37.22433

Bergh, A. (2015). Local educational actors doing of education - a study of how local autonomy meets international and national quality policy rhetoric. *Nordic Journal of Studies in Educational Policy*, *2015*(2). https://doi.org/10.3402/nstep.v1.28146

Bergsmann, E., Klug, J., Burger, C., Först, N., & Spiel, C. (2018). The Competence Screening Questionnaire for Higher Education: Adaptable to the needs of a study programme. *Assessment & Evaluation in Higher Education*, *43*(4), 537–554. https://doi.org/10.1080/02602938.2017.1378617

Bergsmann, E., Klug, J., Schultes, M.-T., Spiel, C., Kelo, M., Ledermüller, K., Blüml, F., Hopbach, A., del Moral, A. B., & Pohl, L. (2016). Handbook for Internal Quality Management in Competence-Based Higher Education (IQM– HE). In.

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

Berk, R. A. (2018). Start spreading the news: Use multiple sources of evidence to evaluate teaching. *The Journal of Faculty Development*, *32*(1), 73–81.

Berk, R. A., Naumann, P. L., & Appling, S. E. (2005). Beyond student ratings: Peer observation of classroom and clinical teaching. *International Journal of Nursing Education Scholarship*, *1*(1).

Biggs, J. (1999). *Teaching for quality learning at university: What the student does*. Society for Research into Higher Education and Open University Press.

Biggs, J. (2003). Aligning teaching and assessing to course objectives. *Teaching and learning in higher education: New trends and innovations*, *2*(April), 13–17.

Blackmore, J. (2009). Academic pedagogies, quality logics and performative universities: Evaluating teaching and what students want. *Studies in Higher Education*, *34*(8), 857–872. https://doi.org/10.1080/03075070902898664

Blaikie, N. (2007). *Approaches to social enquiry: Advancing knowledge* (2nd ed.). Polity.

Blaikie, N., & Priest, J. (2019). *Designing social research: The logic of anticipation*. Polity Press.

Blair, E., & Valdez Noel, K. (2014). Improving higher education practice through student evaluation systems: Is the student voice being heard? *Assessment & Evaluation in Higher Education*, *39*(7), 879–894. https://doi.org/10.1080/02602938.2013.875984

Bleiklie, I. (1998). Justifying the evaluative state: New public management ideals in higher education. *Journal of Public Affairs Education*, *4*(2), 87–100.

Bleiklie, I., & Frølich, N. (2014). Styring, organisering og ledelse i høyere utdanningspolitikk. In *Kvalitet, kapasitet og relevans. Utviklingstrekk i norsk høyere utdanning* (pp. 40–57). Cappelen Damm.

Bloch, C., Degn, L., Nygaard, S., & Haase, S. (2020). Does quality work work? A systematic review of academic literature on quality initiatives in higher education. *Assessment & Evaluation in Higher Education*, 1–18. https://doi.org/10.1080/02602938.2020.1813250

Boote, D. N., & Beile, P. (2005). Scholars before researchers: On the centrality of the dissertation literature review in research preparation. *Educational researcher*, *34*(6), 3–15.

Borch, I. (2020). Lost in translation: From the university's quality assurance system to student evaluation practice. *Nordic Journal of Studies in Educational Policy*, *6*(3), 231–244. https://doi.org/10.1080/20020317.2020.1818447

Borch, I., Sandvoll, R., & Risør, T. (2020). Discrepancies in purposes of student course evaluations: What does it mean to be "satisfied"? *Educational Assessment, Evaluation and Accountability*, *32*(1), 83–102. https://doi.org/10.1007/s11092-020-09315-x

Borch, I., Sandvoll, R., & Risør, T. (2021). Student course evaluation documents: Constituting evaluation practice. *Assessment & Evaluation in Higher Education*. https://doi.org/10.1080/02602938.2021.1899130

Boréus, K., & Bergström, G. (2017). *Analyzing text and discourse: Eight approaches for the social sciences*. SAGE Publications.

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. http://www.pestlhe.org/index.php/pestlhe/article/view/53

Bovill, C., Cook-Sather, A., & Felten, P. (2011). Students as co-creators of teaching approaches, course design, and curricula: Implications for academic developers. *International Journal for Academic Development*, *16*(2), 133–145.

Bowden, J., & Marton, F. (1998). *The university of learning: Beyond quality and competence*. Kogan Page.

Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal*, *9*(2), 27–40.

Boysen, G. A., Kelly, T. J., Raesly, H. N., & Casner, R. W. (2014). The (mis)interpretation of teaching evaluations by college faculty and administrators. *Assessment & Evaluation in Higher Education*, *39*(6), 641–656. https://doi.org/10.1080/02602938.2013.860950

Braga, M., Paccagnella, M., & Pellizzari, M. (2014). Evaluating students' evaluations of professors. *Economics of Education Review*, *41*, 71–88. https://doi.org/10.1016/j.econedurev.2014.04.002

Brandl, K., Mandel, J., & Winegarden, B. (2017). Student evaluation team focus groups increase students' satisfaction with the overall course evaluation process. *Medical Education*, *51*(2), 215–227. https://doi.org/10.1111/medu.13104

Braun, V., & Clarke, V. (2020). One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative Research in Psychology*, 1–25. https://doi.org/10.1080/14780887.2020.1769238

Brinkmann, S., & Kvale, S. (2015). *Interviews: Learning the craft of qualitative research interviewing* (3rd ed.). SAGE Publications.

Brooman, S., Darwent, S., & Pimor, A. (2015). The student voice in higher education curriculum design: Is there value in listening? *Innovations in Education and Teaching International*, *52*(6), 663–674.

Brown, M. J. (2008). Student perceptions of teaching evaluations. *Journal of Instructional Psychology*, *35*(2), 177–181.

Brunsson, N. (2003). Organized hypocrisy. In B. Czarniawska-Joerges & G. Sevon (Eds.), *The northern lights*. Liber.

Burr, V. (1995). *An introduction to social constructionism*. Routledge.

Burr, V. (2015). *Social constructionism* (3rd ed.). Routledge.

Burton, W. B., Civitano, A., & Steiner-Grossman, P. (2012). Online versus paper evaluations: Differences in both quantitative and qualitative data. *Journal of Computing in Higher Education*, *24*(1), 58–69.

Campbell, J. P., & Bozeman, W. C. (2008). The value of student ratings: Perceptions of students, teachers, and administrators. *Community College Journal of Research and Practice*, *32*(1), 13–24. https://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ781786&site=ehost-live

Carbone, A., Ross, B., Phelan, L., Lindsay, K., Drew, S., Stoney, S., & Cottman, C. (2015). Course evaluation matters: Improving students' learning experiences with a peer-assisted teaching programme. *Assessment & Evaluation in Higher Education*, *40*(2), 165–180. https://doi.org/10.1080/02602938.2014.895894

Carrell, S. E., & West, J. E. (2010). Does professor quality matter? Evidence from random assignment of students to professors. *Journal of Political Economy*, *118*(3), 409–432.

Carter, S. M., & Little, M. (2007). Justifying knowledge, justifying method, taking action: Epistemologies, methodologies, and methods in qualitative research. *Qualitative Health Research*, *17*(10), 1316–1328.

Castello, M., & Inesta, A. (2012). Texts as artifacts-in-activity: Developing authorial identity and academic voice in writing academic research papers. In M. Castello & C. Donahue (Eds.), *University writing selves and texts in academic societies* (Vol. 24, pp. 179–200). Emerald. https://doi.org/10.1108/S1572-6304(2012)0000024014

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

Chávez, K., & Mitchell, K. M. W. (2020). Exploring bias in student evaluations: Gender, race, and ethnicity. *PS: Political Science & Politics*, *53*(2), 270–274. https://doi.org/10.1017/S1049096519001744

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

Christensen, G. (2010). Evaluering af evalueringer. *Dansk Universitetspædagogisk Tidsskrift*, *5*(8), 10–16.

Christensen, T., Lægreid, P., & Røvik, K. A. (2020). *Organization theory and the public sector: Instrument, culture and myth* (2nd ed.). Routledge.

Christie, C. A., & Alkin, M. C. (2008). Evaluation theory tree re-examined. *Studies in Educational Evaluation*, *34*(3), 131–135. https://doi.org/10.1016/j.stueduc.2008.07.001

Clayson, D. E. (2009). Student evaluations of teaching: Are they related to what students learn?: A meta-analysis and review of the literature. *31*(1), 16–30. https://doi.org/10.1177/0273475308324086

Cohen, L., Manion, L., & Morrison, K. R. B. (2017). *Research methods in education* (8th ed.). Taylor & Francis Group.

Cook-Sather, A., Krishna Prasad, S., Marquis, E., & Ntem, A. (2019). Mobilizing a culture shift on campus: Underrepresented students as educational developers. *New Directions for Teaching and Learning*, *2019*(159), 21–30. https://doi.org/10.1002/tl.20345

Cooren, F. (2004). Textual agency: How texts do things in organizational settings. *Organization*, *11*(3), 373–393. https://doi.org/10.1177/1350508404041998

Cooren, F. (2015). Textual agency. *The International Encyclopedia of Language and Social Interaction*, 1-5. https://doi.org/https://doi.org/10.1002/9781118611463.wbielsi054

Cordiner, M. (2014). Academic developers as change agents improving quality in a large interprofessional undergraduate subject. *International Journal for Academic Development*, *19*(3), 199–211. https://doi.org/10.1080/1360144X.2013.825618

Cousins, J. B., Goh, S. C., Elliott, C. J., & Bourgeois, I. (2014). Framing the capacity to do and use evaluation. In B. Cousins & I. Bourgeois (Eds.), *Organizational capacity to do and use evaluation* (Vol. 2014, pp. 7–23). John Wiley & Sons. https://doi.org/10.1002/ev.20076

Cousins, J. B., & Leithwood, K. A. (1986). Current empirical research on evaluation utilization. *Review of Educational Research*, *56*(3), 331–364.

Cousins, J. B., & Whitmore, E. (1998). Framing participatory evaluation. *New Directions for Evaluation*, *1998*(80), 5–23.

Creswell, J. W. (2013). *Qualitative inquiry & research design: Choosing among five approaches* (3rd ed.). SAGE Publications.

Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry & research design: Choosing among five approaches* (4th ed.). SAGE Publications.

Cunliffe, A. L. (2004). On becoming a critically reflexive practitioner. *Journal of Management Education*, *28*(4), 407–426. https://doi.org/10.1177/1052562904264440

Cunliffe, A. L. (2016). "On becoming a critically reflexive practitioner" Redux: What does it mean to be reflexive? *Journal of Management Education*, *40*(6), 740–746. https://doi.org/10.1177/1052562916668919

Curby, T., McKnight, P., Alexander, L., & Erchov, S. (2020). Sources of variance in end-of-course student evaluations. *Assessment & Evaluation in Higher Education*, *45*(1), 44–53. https://doi.org/10.1080/02602938.2019.1607249

Curran, S. J. (2020). Pitfalls in using small number statistics in teaching evaluations: A case study. *Assessment & Evaluation in Higher Education*, *45*(3), 419–430. https://doi.org/10.1080/02602938.2019.1664396

Cypress, B. S. (2017). Rigor or reliability and validity in qualitative research: Perspectives, strategies, reconceptualization, and recommendations. *Dimensions of Critical Care Nursing*, *36*(4), 253–263. https://doi.org/10.1097/DCC.0000000000000253

Czarniawska, B. (2008). How to misuse institutions and get away with it: Some reflections on institutional theory (ies). In R. Greenwood, C. Oliver, R. Suddaby, & K. Sahlin (Eds.), *The SAGE handbook of organizational institutionalism* (pp. 769–782). SAGE Publications.

Czarniawska, B., & Joerges, B. (1996). Travel of ideas. In B. Czarniawska & G. Savon (Eds.), *Translating organizational change* (pp. 13–48). Walter de Gruyter.

Dahler-Larsen, P. (2005). Evaluation and public management. In E. Ferlie, L. Lynn, & C. Pollitt (Eds.), *The Oxford handbook of public management* (pp. 615–639). Oxford University Press.

Dahler-Larsen, P. (2006). Evaluation after disenchantment? Five issues shaping the role of evaluation in society. In I. F. Shaw, J. C. Greene, & M. M. Mark (Eds.), *The SAGE Handbook of Evaluation* (pp. 142–160). SAGE Publications. https://doi.org/10.4135/9781848608078

Dahler-Larsen, P. (2011). *The evaluation society*. Stanford University Press.

Dahler-Larsen, P. (2019). *Quality: From Plato to performance* (1st ed.). Palgrave Macmillan.

Dahler-Larsen, P., & Krogstrup, H. K. (2001). Evalueringers konstitutive virkninger. In *Tendenser i evaluering* (pp. 232–245). Syddansk Universitetsforlag.

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*. 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

Datnow, A., & Hubbard, L. (2016). Teacher capacity for and beliefs about data-driven decision making: A literature review of international research. *Journal of Educational Change*, *17*(1), 7–28. https://doi.org/10.1007/s10833-015-9264-2

Davies, M., Hirschberg, J., Lye, J., Johnston, C., & McDonald, I. (2007). Systematic influences on teaching evaluations: The case for caution. *Australian Economic Papers*, *46*(1), 18–38.

Debowski, S. (2014). From agents of change to partners in arms: The emerging academic developer role. *International Journal for Academic Development*, *19*(1), 50–56.

DeLuca, C., Poth, C., & Searle, M. (2009). Evaluation for learning: A cross-case analysis of evaluator strategies. *Studies in Educational Evaluation*, *35*(4), 121–129. https://doi.org/https://doi.org/10.1016/j.stueduc.2009.12.005

Denzin, N. K., & Lincoln, Y. S. (2018). *The SAGE Handbook of Qualitative Research* (5th ed.). SAGE Publications.

Dicker, R., Garcia, M., Kelly, A., & Mulrooney, H. (2019). What does 'quality' in higher education mean? Perceptions of staff, students and employers. *Studies in Higher Education*, *44*(8), 1425–1441.

Dommeyer, C. J., Baum, P., Chapman, K. S., & Hanna, R. W. (2002). Attitudes of business faculty towards two methods of collecting teaching evaluations: Paper vs. online. *Assessment & Evaluation in Higher Education*, *27*(5), 455–462.

Dommeyer, C. J., Baum, P., Hanna, R. W., & Chapman, K. S. (2004). Gathering faculty teaching evaluations by in-class and online surveys: Their effects on response rates and evaluations. *Assessment & Evaluation in Higher Education*, *29*(5), 611–623. https://doi.org/10.1080/02602930410001689171

Donnon, T., Delver, H., & Beran, T. (2010). Student and teaching characteristics related to ratings of instruction in medical sciences graduate programs. *Medical Teacher*, *32*(4), 327–332.

Donovan, J., Mader, C. E., & Shinsky, J. (2010). Constructive student feedback: Online vs. traditional course evaluations. *Journal of Interactive Online Learning*, *9*(3), 283–296.

Dunegan, K. J., & Hrivnak, M. W. (2003). Characteristics of mindless teaching evaluations and the moderating effects of image compatibility. *Journal of Management Education*, *27*(3), 280–303. https://doi.org/10.1177/1052562903027003002

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

El Ansari, W., & Moseley, L. (2011). You get what you measure: Assessing and reporting student satisfaction with their health and social care educational experience. *Nurse Education Today*, *31*(2), 173–178. https://doi.org/10.1016/j.nedt.2010.05.014

El Hassan, K. (2009). Investigating substantive and consequential validity of student ratings of instruction. *Higher Education Research & Development*, *28*(3), 319–333. https://doi.org/10.1080/07294360902839917

Elken, M., Maassen, P., Nerland, M., Prøitz, T. S., Stensaker, B., & Vabø, A. (2020). *Quality work in higher education: Organisational and pedagogical dimensions*. Springer International Publishing. https://doi.org/10.1007/978-3-030-41757-4

Elken, M., & Stensaker, B. (2018). Conceptualising 'quality work' in higher education. *Quality in Higher Education*, *24*(3), 189–202.

Elken, M., & Stensaker, B. (2020a). Quality work: Reflections and refinements. In M. Elken, P. Maassen, M. Nerland, T. Prøitz, B. Stensaker, & A. Vabø (Eds.), *Quality work in higher education: Organisational and pedagogical dimensions* (pp. 175–186). Springer International Publishing. https://doi.org/10.1007/978-3-030-41757-4_10

Elken, M., & Stensaker, B. (2020b). Researching 'quality work' in higher education. In M. Elken, P. Maassen, T. Prøitz, B. Stensaker, & A. Vabø (Eds.), *Quality work in higher education: Organisational and pedagogical dimensions* (pp. 1–17). Springer International Publishing. https://doi.org/10.1007/978-3-030-41757-4_1

ENQA. (2015). *Standards and Guidelines for Quality Assurance in the European higher education area (ESG)*. Brussels, Belgium Retrieved from http://www.enqa.eu/index.php/home/esg/

Eriksson-Zetterquist, U. (2009). *Institutionell teori: idéer, moden, förändring*. Liber.

Ernst, D. (2014). Expectancy theory outcomes and student evaluations of teaching. *Educational Research and Evaluation*, *20*(7-8), 536–556.

Ese, J. (2019). *Defending the university?: Academics' reactions to managerialism in Norwegian higher education* [Ph.D., Karlstads universitet]. Karlstad.

Esmark, A., Bagge Laustsen, C., & Åkerstrøm Andersen, N. (2005). *Socialkonstruktivistiske analysestrategier*. Roskilde Universitetsforlag.

Estelamani, H. (2015). The effects of survey timing on student evaluation of teaching measures obtained using online surveys. *Journal of Marketing Education*, *37*(1), 54–64.

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. *PloS one*, *14*(2), e0209749. https://doi.org/10.1371/journal.pone.0209749

Feldman, K. A. (2007). Identifying exemplary teachers and teaching: Evidence from student ratings[1]. In R. P. Perry & J. C. Smart (Eds.), *The Scholarship of Teaching and Learning in higher education: An evidence-based perspective* (pp. 93–143). Springer. https://doi.org/10.1007/1-4020-5742-3_5

Felten, P., Abbot, S., Kirkwood, J., Long, A., Lubicz-Nawrocka, T., Mercer-Mapstone, L., & Verwoord, R. (2019). Reimagining the place of students in academic development. *International Journal for Academic Development*, *24*(2), 192–203.

Fischer, E., & Hänze, M. (2019). Bias hypotheses under scrutiny: Investigating the validity of student assessment of university teaching by means of external observer ratings. *Assessment & Evaluation in Higher Education*, *44*(5), 772–786. https://doi.org/10.1080/02602938.2018.1535647

Fitzpatrick, J. L. (2012). Commentary –Collaborative evaluation within the larger evaluation context. *Evaluation and program planning*, *35*(4), 558. https://doi.org/10.1016/j.evalprogplan.2011.12.012

Fleischer, D. N., & Christie, C. A. (2009). Evaluation use: Results from a survey of U.S. American Evaluation Association members. *American Journal of Evaluation*, *30*(2), 158–175. https://doi.org/10.1177/1098214008331009

Franklin, J. (2001). Interpreting the numbers: Using a narrative to help others read student evaluations of your teaching accurately. *New Directions for Teaching and Learning*, *87*, 85–100.

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

Froestad, W., & Haakstad, J. (2009). *Å rapportere kvalitet: en studie av rapporter til styret om utdanningskvalitet og kvalitetsarbeid for årene 2005, 2006 og 2007 i 18 institusjoner*. NOKUT.

Furu, E. M., Skrøvset, S., & Slettbakk, Å. (2020). Lærere i nye roller som oversettere: Hvilket handlingsrom har de for utvikling? *Nordisk tidsskrift for utdanning og praksis*, *14*(3), 1–16. https://doi.org/10.23865/up.v14.2206

Gakhal, S., & Wilson, C. (2019). Is students' qualitative feedback changing, now it is online? *Assessment & Evaluation in Higher Education*, *44*(3), 476–488.

Galbraith, C., Merrill, G., & Kline, D. (2012). Are student evaluations of teaching effectiveness valid for measuring student learning outcomes in business related classes? A neural network and Bayesian analyses. *Journal of the Association for Institutional Research*, *53*(3), 353–374. https://doi.org/10.1007/s11162-011-9229-0

Gibbs, G. (2013). Reflections on the changing nature of educational development. *International Journal for Academic Development*, *18*(1), 4–14.

Giles, A., Martin, Sylvia, Bryce, Deborah & Hendry Graham. (2004). Students as partners in evaluation: Student and teacher perspectives. *Assessement & Evaluation in Higher Education*, *29*(6), 682–685. https://doi.org/1010800260293042000227227

Ginns, P., & Barrie, S. (2009). Developing and testing a student-focused teaching evaluation survey for university instructors. *Psychological Reports*, *104*(3), 1019–1032. https://doi.org/10.2466/PR0.104.3.1019-1032

Glosvik, Ø. (2017). Eit translasjonsteoretisk perspektiv på kunnskapsleiing i skular. In *Immateriell Kapital* (pp. 240–257). https://doi.org/10.18261/9788215028163-2017-13

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

Gornitzka, A., & Larsen, I. M. (2004). Towards professionalisation? Restructuring of administrative work force in universities. *Higher Education: The International Journal of Higher Education and Educational Planning*, *47*(4), 455–471. https://doi.org/10.1023/B:HIGH.0000020870.06667.f1

Gornitzka, Å., Stensaker, B., Smeby, J. C., & De Boer, H. (2004). Contract arrangements in the Nordic countries—Solving the efficiency/effectiveness dilemma? *Higher Education in Europe*, *29*(1), 87–101.

Grant, M. J., & Booth, A. (2009). A typology of reviews: An analysis of 14 review types and associated methodologies. *Health Information & Libraries Journal*, *26*(2), 91–108.

Grimes, A., Medway, D., Foos, A., & Goatman, A. (2017). Impact bias in student evaluations of higher education. *Studies in Higher Education*, *42*(6), 945–962. https://doi.org/10.1080/03075079.2015.1071345

Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation*. SAGE Publications.

Gulbrandsen, M., & Stensaker, B. (2003). Evaluering av forskning og utdanning–bidrag til en helhetlig kvalitetsutvikling? In I. M. Larsen & B. Stensaker (Eds.), *Tradisjon og tilpasning, organisering og styring av universitetene* (pp. 79–98). Cappelen Aakademisk Forlag.

Gynnild, V. (2002). Studentevaluering av undervisning. *Norsk pedagogisk tidsskrift*(01), 25–38.

Hadad, Y., Keren, B., & Naveh, G. (2020). The relative importance of teaching evaluation criteria from the points of view of students and faculty. *Assessment & Evaluation in Higher Education*, *45*(3), 447–459. https://doi.org/10.1080/02602938.2019.1665623

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

Hamermesh, D. S., & Parker, A. (2005). Beauty in the classroom: Instructors' pulchritude and putative pedagogical productivity. *Economics of Education Review*, *24*(4), 369–376. https://doi.org/https://doi.org/10.1016/j.econedurev.2004.07.013

Hammer, R., Peer, E., & Babad, E. (2018). Faculty attitudes about student evaluations and their relations to self-image as teacher. *Social Psychology of Education*, *21*(3), 517–537.

Hammersley, M. (2007). The issue of quality in qualitative research. *International journal of research & method in education*, *30*(3), 287–305. https://doi.org/10.1080/17437270701614782

Hampton, S. E., & Reiser, R. A. (2004). Effects of a theory-based feedback and consultation process on instruction and learning in college classrooms. *Research in Higher Education*, *45*(5), 497–527. https://doi.org/10.1023/b:rihe.0000032326.00426.d5

Hanken, I. M. (2011). Student evaluation of teaching from the actors' perspective. *Quality in Higher Education*, *17*(2), 245–256. https://doi.org/10.1080/13538322.2011.582797

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

Harris, M. J. (2017). *Evaluating public and community health programs* (2nd ed.). John Wiley & Sons.

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

Harvey, L. (2003). Student feedback 1. *Quality in Higher Education*, *9*(1), 3–20. https://doi.org/10.1080/13538320308164

Harvey, L. (2004-20). Analytic Quality Glossary. In: Quality Research International.

Harvey, L. (2007). Epistemology of quality. *Perspectives in education*, *25*(3), 1–13.

Harvey, L., & Green, D. (1993). Defining quality. *Assessment & Evaluation in Higher Education*, *18*(1), 9–34. https://doi.org/10.1080/0260293930180102

Harvey, L., & Newton, J. (2004). Transforming quality evaluation. *Quality in Higher Education*, *10*(2), 149-165. https://doi.org/10.1080/1353832042000230635

Harvey, L., & Williams, J. (2010). Fifteen years of quality in higher education. *Quality in Higher Education*, *16*(1), 3–36. https://doi.org/10.1080/13538321003679457

Hauer, K. E., & Papadakis, M. A. (2010). Assessment of the contributions of clinician educators. *Journal of general internal medicine*, *25*(1), 5–6. https://doi.org/10.1007/s11606-009-1186-3

Havnes, A., & Stensaker, B. (2006). Educational development centres: From educational to organisational development? *Quality Assurance in Education*, *14*(1), 7–20. https://doi.org/10.1108/09684880610643584

Heine, P., & Maddox, N. (2009). Student perceptions of the faculty course evaluation process: An exploratory study of gender and class differences. *Research in Higher Education Journal*, *3*, 1–10.

Henry, G. T., & Mark, M. M. (2003). Beyond use: Understanding evaluation's influence on attitudes and actions. *American Journal of Evaluation*, *24*(3), 293–314. https://doi.org/10.1177/109821400302400302

Hessler, M., Pöpping, D. M., Hollstein, H., Ohlenburg, H., Arnemann, P. H., Massoth, C., Seidel, L. M., Zarbock, A., & Wenk, M. (2018). Availability of cookies during an academic course session affects evaluation of teaching. *Medical Education*, *52*(10), 1064–1072. https://doi.org/10.1111/medu.13627

Hobson, S. M., & Talbot, D. M. (2001). Understanding student evaluations: What all faculty should know. *College Teaching*, *49*(1), 26–31.

Hoel, A., & Dahl, T. I. (2018). Why bother? Student motivation to participate in student evaluations of teaching. *Assessement & Evaluation in Higher Education*, 361–378.

Højlund, S. (2014). Evaluation use in the organizational context–Changing focus to improve theory. *Evaluation*, *20*(1), 26–43.

Hulpiau, V., Masschelein, E., Der Stockt, L. V., Verhesschen, P., & Waeytens, K. (2007). A system of student feedback: Considerations of academic staff taken into account. *Tertiary Education and Management*, *13*(1), 35–45. https://doi.org/10.1080/13583880601145520

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

Huxham, M., Scoles, J., Green, U., Purves, S., Welsh, Z., & Gray, A. (2017). "Observation has set in": Comparing students and peers as reviewers of teaching. *Assessment & Evaluation in Higher Education*, *42*(6), 887–899. https://doi.org/10.1080/02602938.2016.1204594

Järvinen, M., & Bertilsson, M. (1998). *Socialkonstruktivisme: bidrag til en kritisk diskussion*. Hans Reitzels Forlag.

Järvinen, M., & Mik-Meyer, N. (2017). *Kvalitativ analyse: syv traditioner*. Hans Reitzels Forlag.

Johannessen, I. L. (2020). *Ledelse og kvalitetsutvikling i kommunale barnehager -En kvalitativ studie om lederes igangsetting, oversetting og iverksetting av*

kvalitetsutviklingsarbeid i barnehagen [Ph.D, UiT Norges arktiske universitet]. Tromsø.

Johansen, M. B. (2020). *Studieprogramledelse i høyere utdanning – i spenningsfelt mellom struktur og handlingsrom* [Ph.D, NTNU]. Trondheim.

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.

Justesen, L., & Mik-Meyer, N. (2010). *Kvalitative metoder i organisations- og ledelsesstudier*. Hans Reitzels Forlag.

Kalpazidou Schmidt, E. (2017). Quality assurance policies and practices in Scandinavian higher education systems: Convergence or different paths? *Journal of Higher Education Policy & Management*, *39*(3), 247–265. https://doi.org/10.1080/1360080X.2017.1298194

Kellaghan, T., & Stufflebeam, D. L. (2012). *International handbook of educational evaluation: Part one: Perspectives/Part two: Practice* (Vol. 9). Springer Science & Business Media.

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

Kember, D., & Wong, A. (2000). Implications for evaluation from a study of students' perceptions of good and poor teaching. *Higher Education*, *40*(1), 69–97.

King, J. A., & Alkin, M. C. (2019). The centrality of use: Theories of evaluation use and influence and thoughts on the first 50 years of use research. *American Journal of Evaluation*, *40*(3), 431–458.

King, J. A., & Stevahn, L. (2013). *Interactive evaluation practice: Mastering the interpersonal dynamics of program evaluation*. SAGE Publications.

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

Kirkhart, K. E. (2000). Reconceptualizing evaluation use: An integrated theory of influence. *New Directions for Evaluation*, *2000*(88), 5–23. https://doi.org/doi:10.1002/ev.1188

Kite, M. E., Subedi, P. C., & Bryant-Lees, K. B. (2015). Students' perceptions of the teaching evaluation process. *Teaching of Psychology*, *42*(4), 307–314. https://doi.org/10.1177/0098628315603062

Kitzinger, J., & Barbour, R. (1999). *Developing focus group research: Politics, theory and practice*. SAGE Publications.

Klemenčič, M., & Chirikov, I. (2015). How do we know how students experience higher education? On the use of student surveys. In *The European higher education area* (pp. 361–379). Springer.

Knapper, C. (2001). Broadening our approach to teaching evaluation. *New Directions for Teaching, Learning and instruction*, *2001*(88), 3–9.

Kogan, J. R., & Shea, J. A. (2007). Course evaluation in medical education. *Teaching and Teacher Education*, *23*(3), 251–264. https://doi.org/10.1016/j.tate.2006.12.020

Kogan, L. R., Schoenfeld-Tacher, R., & Hellyer, P. W. (2010). Student evaluations of teaching: Perceptions of faculty based on gender, position, and rank. *Teaching in Higher Education*, *15*(6), 623–636.

Krogstrup, H. K., & Dahler-Larsen, P. (2000). Evaluering og institutionelle standarder: nyinstitutionelle betragtninger af evaluering som vor tids ritual. *Politica*, *32*(3), 45–61.

Krueger, R. A., & Casey, M. A. (2015a). Focus group interviewing. In K. E. Newcomer, H. P. Hatry, & J. S. Wholey (Eds.), *Handbook of practical program evaluation* (4th ed., pp. 506–534). John Wiley & Sons. https://doi.org/10.1002/9781119171386.ch20

Krueger, R. A., & Casey, M. A. (2015b). *Focus groups: A practical guide for applied research* (5th ed.). SAGE Publications.

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

Kunnskapsdepartementet. (2017). *Kultur for kvalitet i høyere utdanning Meld. St. 16 (2016-2017)*. Oslo: Kunnskapsdepartement Retrieved from https://www.regjeringen.no/no/dokumenter/meld.-st.-16-20162017/id2536007/

Kuper, A., Lingard, L., & Levinson, W. (2008). Critically appraising qualitative research. *BMJ*, *337*(73). https://doi.org/10.1136/bmj.a1035

Kuper, A., Whitehead, C., & Hodges, B. D. (2013). Looking back to move forward: Using history, discourse and text in medical education research: AMEE Guide No. 73. *Medical Teacher*, *35*(1), 849–860. https://doi.org/10.3109/0142159X.2012.748887

Lang, J. W., & Kersting, M. (2007). Regular feedback from student ratings of instruction: Do college teachers improve their ratings in the long run? *Instructional Science*, *35*(3), 187–205.

Langbein, L. (2008). Management by results: Student evaluation of faculty teaching and the mis-measurement of performance. *Economics of Education Review*, *27*(4), 417–428.

Larsen, H. (2020, 03.11.2020). Kritikere krever endringer i Studiebarometeret. *Khrono*. https://khrono.no/kritikere-krever-endringer-i-studiebarometeret/528746

Leckey, J., & Neill, N. (2001). Quantifying quality: The importance of student feedback. *Quality in Higher Education*, *7*(1), 19–32. https://doi.org/10.1080/13538320120045058

Lemos, M., Queirós, C., Teixeira, P., & Menezes, I. (2011). Development and validation of a theoretically based, multidimensional questionnaire of student evaluation of university teaching. *Assessment & Evaluation in Higher Education*, *36*(7), 843–864.

Leth Andersen, H., Ravn, J., & Colerick, T. (2013). Undervisningsevalueringens dilemmaer: kontrol og udvikling, to værdener. *Dansk Universitetspædagogisk Tidsskrift*, *8*(15), 85–98.

Leung, L. (2015). Validity, reliability, and generalizability in qualitative research. *Journal of Family Medicine and Primary Care*, *4*(3), 324–327. https://doi.org/10.4103/2249-4863.161306

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

Lillejord, S., Elstad, E., & Kavli, H. (2018). Teacher evaluation as a wicked policy problem. *Assessment in Education: Principles, Policy & Practice*, *25*(3), 291–309. https://doi.org/10.1080/0969594X.2018.1429388

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. SAGE Publications.

Lincoln, Y. S., & Guba, E. G. (2013). *The constructivist credo*. Taylor & Francis Group.

Lindgren, B.-M., Lundman, B., & Graneheim, U. H. (2020). Abstraction and interpretation during the qualitative content analysis process. *International Journal of Nursing Studies*, *108*, 103632-103632. https://doi.org/10.1016/j.ijnurstu.2020.103632

Linse, A. R. (2017). Interpreting and using student ratings data: Guidance for faculty serving as administrators and on evaluation committees. *Studies in Educational Evaluation*, *54*, 94–106. https://doi.org/https://doi.org/10.1016/j.stueduc.2016.12.004

Lipsey, N., & Shepperd, J. (2020). Examining strategies to increase student evaluation of teaching completion rates. *Assessment & Evaluation in Higher Education*, 1–14. https://doi.org/10.1080/02602938.2020.1782343

Little, B., & Williams, R. (2010). Students' roles in maintaining quality and in enhancing learning: Is there a tension? *Quality in Higher Education*, *16*(2), 115–127.

Logermann, F., & Leišytė, L. (2015). Students as stakeholders in the policy context of the European Standards and Guidelines for quality assurance in higher education institutions. In A. Curaj, L. Matei, R. Pricopie, J. Salmi, & P. Scott (Eds.), *The European higher education area: Between critical reflections and future policies* (pp. 685–701). Springer International Publishing. https://doi.org/10.1007/978-3-319-20877-0_43

Malterud, K. (2001). Qualitative research: Standards, challenges, and guidelines. *The Lancet*, *358*(9280), 483–488.

Malterud, K. (2015). Theory and interpretation in qualitative studies from general practice: Why and how? *Scandinavian Journal of Public Health*, *44*(2), 120–129. https://doi.org/10.1177/1403494815621181

Mark, E. (2013). Student satisfaction and the customer focus in higher education. *Journal of Higher Education Policy and Management*, *35*(1), 2–10. https://doi.org/10.1080/1360080X.2012.727703

Mark, M. M. (2009). *Evaluation, method choices and pathways to consequences: Trying to make sense of how evaluation can contribute to sensemaking*. SAGE Publications.

Mark, M. M., & Henry, G. T. (2004). The mechanisms and outcomes of evaluation influence. *Evaluation*, *10*(1), 35–57.

Marsh, H. W. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. *International Journal of Educational Research*, *11*(3), 253–388.

Marsh, H. W. (2007). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases and usefulness. In *The Scholarship of Teaching and Learning in higher education: An evidence-based perspective* (pp. 319–383). Springer.

Marsh, H. W., & Roche, L. A. (2000). Effects of grading leniency and low workload on students' evaluations of teaching: Popular myth, bias, validity, or innocent Bystanders? *Journal of Educational Psychology*, *92*(1), 202.

Mau, R. R., & Opengart, R. (2012). Comparing ratings: In-class (paper) vs. out of class (online) student evaluations. *Higher Education Studies*, *2*(3), 55.

Maxwell, J. A. (2006). Literature reviews of, and for, educational research: A commentary on Boote and Beile's "Scholars before researchers". *Educational researcher*, *35*(9), 28–31.

Maxwell, J. A. (2013). *Qualitative research design: An interactive approach* (3rd ed., Vol. 41). SAGE Publications.

McClain, L., Gulbis, A., & Hays, D. (2018). Honesty on student evaluations of teaching: effectiveness, purpose, and timing matter! *Assessment & Evaluation in Higher Education*, *43*(3), 369–385. https://doi.org/10.1080/02602938.2017.1350828

McCulloch, A. (2009). The student as co-producer: Learning from public administration about the student–university relationship. *Studies in Higher Education*, *34*(2), 171–183.

McGrath, C. (2020). Academic developers as brokers of change: Insights from a research project on change practice and agency. *International Journal for Academic Development*, *25*(2), 94–106. https://doi.org/10.1080/1360144X.2019.1665524

McPherson, M. A. (2006). Determinants of how students evaluate teachers. *The Journal of Economic Education*, *37*(1), 3–20.

Mengel, F., Sauermann, J., & Zölitz, U. (2019). Gender bias in teaching evaluations. *Journal of the European Economic Association*, *17*(2), 535–566.

Mercer, J. (2007). The challenges of insider research in educational institutions: Wielding a double-edged sword and resolving delicate dilemmas. *Oxford Review of Education*, *33*(1), 1–17.

Mercer-Mapstone, L., Dvorakova, S. L., Matthews, K. E., Abbot, S., Cheng, B., Felten, P., Knorr, K., Marquis, E., Shammas, R., & Swaim, K. (2017). A systematic literature review of students as partners in higher education. *International Journal for Students as Partners*, *1*(1), 2541–2557.

Merriam, S. B., & Tisdell, E. J. (2016). *Qualitative research: A guide to design and implementation* (4th ed.). John Wiley & Sons.

Meyer, J. W., & Rowan, B. (1977). Institutionalized organizations: Formal structure as myth and ceremony. *American Journal of Sociology*, *83*(2), 340–363.

Michelsen, S., & Aamodt, P. O. (2007). *Evaluering av Kvalitetsreformen. Sluttrapport.* Norges Forskningsråd. http://hdl.handle.net/11250/279245

Mitry, D. J., & Smith, D. E. (2014). Student evaluations of faculty members: A call for analytical prudence. *Journal on Excellence in College Teaching*, *25*(2), 56–67. https://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ1041380&site=ehost-live

Mjøs, O. D., & Utvalget for høgre utdanning. (2000). *Frihet med ansvar: om høgre utdanning og forskning i Norge: utredning fra et utvalg oppnevnt ved kongelig resolusjon 30. April 1998: Avgitt til Kirke-, utdannings- og forskningsdepartementet 8. Mai 2000* (8258305255). https://www.regjeringen.no/no/dokumenter/nou-2000-14/id142780/

Moldt, C. (2006). Fra meningsløse til meningsfulde performancemålinger–en case om kvalitetsmåling i undervisningssektoren. *Økonomistyring og Informatik*, *22*(5).

Moore, S., & Kuol, N. (2005). Students evaluating teachers: Exploring the importance of faculty reaction to feedback on teaching. *Teaching in Higher Education*, *10*(1), 57–73.

Morgan, D. A., Sneed, J., & Swinney, L. (2003). Are student evaluations a valid measure of teaching effectiveness: Perceptions of accounting faculty members and administrators. *Management Research News*, *26*(7), 17–32.

Moskal, A. C. M., Stein, S. J., & Golding, C. (2016). Can you increase teacher engagement with evaluation simply by improving the evaluation system? *Assessment & Evaluation in Higher Education*, *41*(2), 1–15. https://doi.org/10.1080/02602938.2015.1007838

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

Nasser-Abu Alhija, F. (2017). Teaching in higher education: Good teaching through students' lens. *Studies in Educational Evaluation*, *54*, 4–12. https://doi.org/10.1016/j.stueduc.2016.10.006

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/https://doi.org/10.1080/713692740

Newton, J. (2002). Views from below: Academics coping with quality. *Quality in Higher Education*, *8*(1), 39–61. https://doi.org/10.1080/13538320220127434

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. In. SAGE Publications. https://doi.org/10.4135/9781452226606

NOKUT. *Studiebarometeret*. NOKUT. Retrieved 17/03/21 from https://studiebarometeret.no/en/

NOKUT. (2012). *Evaluering av system for kvalitetssikring av utdanningen ved Universitetet i Tromsø* (NOKUTs tilsynsrapporter). Retrieved from

https://www.nokut.no/contentassets/c3c6b0f301954d81a099dcf4de2f3531/universitete t_i_tromso_ks_2012.pdf

NOKUT. (2016). *Kvalitetsområder for studieprogram, Veiledende dokument, 12. Desember 2016* (NOKUTs veiledninger).

Norges Forskningsråd. *Prosjektbanken Forskningsrådet prosjekt SLUSEN Nr 260359*. Retrieved 17/03/21 from https://prosjektbanken.forskningsradet.no/explore/projects?Kilde=FORISS&distributi on=Ar&chart=bar&calcType=funding&Sprak=no&sortBy=score&sortOrder=desc&re sultCount=30&offset=0&Fritekst=260359

Nørholm, M. (2008). *Om den sociale og symbolske funktion af uddannelsesevalueringer: Konstruktionen af et objekt* [Ph.D, Aarhus universitet]. Aarhus.

Norwegian Centre for Research Data. *Data protection services*. Retrieved 21/03/21 from https://www.nsd.no/en/

Norwegian National Research Ethics Committee. *Guidelines*. Retrieved 17/03/21 from https://www.forskningsetikk.no/en/guidelines/

NOU. (2015). *Produktivitet–grunnlag for vekst og velferd–produktivitetskommisjonens første rapport*.

Nulty, D. D. (2008). The adequacy of response rates to online and paper surveys: What can be done? *Assessment & Evaluation in Higher Education*, *33*(3), 301–314.

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

O'Neill, P. A., Willis, S. C., & Jones, A. (2002). A model of how students link Problem-Based Learning with clinical experience through "elaboration". *Academic Medicine*, *77*(6), 552–561.

O'Sullivan, R. G. (2004). *Practicing evaluation: A collaborative approach*. SAGE Publications.

Oliveira, C. G., Oliveira, P. C., & Costa, N. (2012). Students' and teachers' perspectives about quality of engineering education in Portugal. *European Journal of Engineering Education*, *37*(1), 49–57. https://doi.org/10.1080/03043797.2011.653551

Ottoson, J. M. (2009). Knowledge-for-action theories in evaluation: Knowledge utilization, diffusion, implementation, transfer, and translation. *New Directions for Evaluation*, *2009*(124), 7–20.

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.). SAGE Publications.

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

Patton, M. Q. (2007). Process use as a usefulism. *New Directions for Evaluation*, *2007*(116), 99–112.

Patton, M. Q. (2008). *Utilization-Focused Evaluation* (4th ed.). SAGE Publications.

Patton, M. Q. (2011). *Essentials of Utilization-Focused Evaluation*. SAGE Publications.

Patton, M. Q. (2015). *Qualitative research & evaluation methods: Integrating theory and practice* (4th ed.). SAGE Publications.

Patton, M. Q. (2018). *Facilitating evaluation: Principles in practice*. SAGE Publications.

Penny, A. R. (2003). Changing the agenda for research into students' views about university teaching: Four shortcomings of SRT research. *Teaching in Higher Education*, *8*(3), 399–411. https://doi.org/10.1080/13562510309396

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

Peräkylä, A., & Ruusuvuori, J. (2011). Analyzing talk and text. In *The SAGE Handbook of Qualitative Research* (4th ed., pp. 351–374). SAGE Publications.

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

Piccinin, S., & Moore, J.-P. (2002). The impact of individual consultation on the teaching of younger versus older faculty. *The International Journal for Academic Development*, *7*(2), 123–134.

Polit, D. F., & Beck, C. T. (2020). *Nursing research: Generating and assessing evidence for nursing practice* (11th ed.). Lippincott Williams & Wilkins.

Porter, S. R., & Umbach, P. D. (2006). Student survey response rates across institutions: Why do they vary? *Research in Higher Education*, *47*(2), 229–247.

Porter, S. R., & Whitcomb, M. E. (2005). Non-response in student surveys: The role of demographics, engagement and personality. *Research in Higher Education*, *46*(2), 127–152.

Preskill, H., & Boyle, S. (2008). A multidisciplinary model of Evaluation Capacity Building. *American Journal of Evaluation*, *29*(4), 443–459.

Preskill, H., & Torres, R. T. (2000). The learning dimension of evaluation use. *New Directions for Evaluation*, *2000*(88), 25–37.

Prior, L. (2003). *Using documents in social research*. SAGE Publications.

Prior, L. (2008). Repositioning documents in social research. *Sociology*, *42*(5), 821–836.

Prosser, M., Ramsden, P., Trigwell, K., & Martin, E. (2003). Dissonance in experience of teaching and its relation to the quality of student learning. *Studies in Higher Education*, *28*(1), 37–48. https://doi.org/10.1080/03075070309299

Putnam, L., & Nicotera, A. M. (2009). *Building theories of organization: The constitutive role of communication*. Routledge.

Rabiee, F. (2004). Focus-group interview and data analysis. *Proceedings of the Nutrition Society*, *63*(4), 655–660. https://doi.org/10.1079/PNS2004399

Ramsden, P. (2003). *Learning to teach in higher education*. Routledge.

Randolph, J. (2009). A guide to writing the dissertation literature review. *Practical Assessment, Research, and Evaluation*, *14*(1), 13.

Reisenwitz, T. H. (2016). Student evaluation of teaching: An investigation of nonresponse bias in an online context. *Journal of Marketing Education*, *38*(1), 7–17. https://doi.org/10.1177/0273475315596778

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

Rolfe, G. (2006). Validity, trustworthiness and rigour: Quality and the idea of qualitative research. *Journal of Advanced Nursing*, *53*(3), 304–310. https://doi.org/10.1111/j.1365-2648.2006.03727.x

Røvik, K. A. (1998). *Moderne organisasjoner: Trender i organisasjonstenkningen ved tusenårsskiftet*. Fagbokforlaget.

Røvik, K. A. (2007). *Trender og translasjoner: ideer som former det 21. århundrets organisasjon*. Universitetsforlaget.

Røvik, K. A. (2011). From fashion to virus: An alternative theory of organizations' handling of management ideas. *Organization Studies*, *32*(5), 631–653. https://doi.org/10.1177/0170840611405426

Røvik, K. A. (2014). Translasjon, -en alternativ doktrine for implementering. In K. A. Røvik, T. V. Eilertsen, & E. M. Furu (Eds.), *Reformideer i norsk skole, spredning, oversettelse og implementering* (pp. 403–417). Cappelen Damm Akademisk.

Røvik, K. A. (2016). Knowledge transfer as translation: Review and elements of an instrumental theory. *International Journal of Management Reviews*, *18*(3), 290–310. https://doi.org/10.1111/ijmr.12097

Røvik, K. A. (2019). Instrumental understanding of management ideas. In A. Sturdy, S. Heusinkveld, T. Reay, & D. Strang (Eds.), *The Oxford handbook of management ideas* (pp. 121–137). Oxford University Press.

Roxå, T., Ahmad, A., Barrington, J., Van Maaren, J., & Cassidy, R. (2021). Reconceptualising student ratings of teaching to support quality discourse on student learning: A systems perspective. *Higher Education*. https://doi.org/https://doi.org/10.1007/s10734-020-00615-1

Roxå, T., & Bergström, M. (2013). Kursvärderingar i system–akademiska lärares upplevelser och organisationens förmåga till utveckling. *Högre utbildning*, *3*(3), 225–236.

Ryan, K., & Cousins, J. B. (2009). *The SAGE international handbook of educational evaluation*. SAGE Publications.

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

Safavi, S., Bakar, K., Tarmizi, R., & Alwi, N. (2013). Faculty perception of improvements to instructional practices in response to student ratings. *Educational Assessment, Evaluation and Accountability*, *25*(2), 143–153. https://doi.org/10.1007/s11092-013-9160-3

Saldaña, J. (2013). *The coding manual for qualitative researchers* (2nd ed.). SAGE Publications.

Sanderson, I. (2000). Evaluation in complex policy systems. *Evaluation*, *6*(4), 433–454.

Saunders, D. B. (2014). Exploring a customer orientation: Free-market logic and college students. *The Review of Higher Education*, *37*(2), 197–219.

Saunders, M. (2011). Setting the scene: The four domains of evaluative practice in higher education. In M. Saunders, P. Trowler, & R. Bamber (Eds.), *Reconceptualizing evaluation in higher education: The practice turn* (pp. 1–17). ProQuest Ebook Central. https://ebookcentral.proquest.com/lib/tromsoub-ebooks/detail.action?docID=714184

Saunders, M., Trowler, P., Bamber, V., & Trowler, P. (2011). *Reconceptualising evaluation in higher education: The practice turn*. McGraw-Hill Education. http://ebookcentral.proquest.com/lib/tromsoub-ebooks/detail.action?docID=714184

Saunders, M. N. K., & Williams, C. (2005). From evaluation towards an agenda for quality improvement: The development and application of the template process. *Active Learning in Higher Education*, *6*(1), 60–72. https://doi.org/10.1177/1469787405049944

Sax, L. J., Gilmartin, S. K., & Bryant, A. N. (2003). Assessing response rates and nonresponse bias in web and paper surveys. *Research in Higher Education*, *44*(4), 409–432.

Sax, L. J., Gilmartin, S. K., Lee, J. J., & Hagedorn, L. S. (2008). Using web surveys to reach community college students: An analysis of response rates and response bias. *Community College Journal of Research and Practice*, *32*(9), 712–729.

Schindler, L. A., Puls Elvidge, S., Welzant, H., & Crawford, L. (2015). Definitions of quality in higher education: A synthesis of the literature. *Higher Learning Research Communications*, *5*(3), 3–13. https://doi.org/10.18870/hlrc.v5i3.244

Schuck, S., Gordon, S., & Buchanan, J. (2008). What are we missing here? Problematising wisdoms on teaching quality and professionalism in higher education. *Teaching in Higher Education*, *13*(5), 537–547.

Schwandt, T. A. (2009). Globalizing influences on the western evaluation imaginary. In K. Ryan & B. Cousins (Eds.), *The SAGE international handbook of educational evaluation* (pp. 19-36). SAGE Publications.

Scott, J. (1990). *A matter of record: Documentary sources in social research*. Polity Press.

Scott, W. R. (2014). *Institutions and organizations: Ideas, interests, and identities* (4th ed.). SAGE Publications.

Scriven, M. (1967). The methodology of evaluation. In R. Tyler, R. Gagne, & M. Scriven (Eds.), *Perspectives of curriculum evaluation*. Rand McNally.

Scriven, M. (1991). *Evaluation thesaurus* (4th ed.). SAGE Publications.

Scriven, M. (1996). Types of Evaluation and Types of Evaluator. *Evaluation Practice*, *17*(2), 151–161. https://doi.org/10.1177/109821409601700207

Shah, M., Cheng, M., & Fitzgerald, R. (2017). Closing the loop on student feedback: The case of Australian and Scottish universities. *Higher Education*, *74*(1), 115–129.

Shao, L. P., Anderson, L. P., & Newsome, M. (2007). Evaluating teaching effectiveness: where we are and where we should be. *Assessment & Evaluation in Higher Education*, *32*(3), 355–371. https://doi.org/10.1080/02602930600801886

Sharrock, G. (2000). Why students are not (just) customers (and other reflections on life after George). *Journal of Higher Education Policy and Management*, *22*(2), 149–164. https://doi.org/10.1080/713678141

Shulha, L. M., & Cousins, J. B. (1997). Evaluation use: Theory, research, and practice since 1986. *Evaluation Practice*, *18*(3), 195–208. https://doi.org/10.1016/S0886-1633(97)90027-1

Silverman, D. (2013). *Doing qualitative research* (4th ed.). SAGE Publications.

Siu, K. W. M. (2012). Formative student feedback: Enhancing the quality of learning and teaching. In C. S. Nair (Ed.), *Enhancing learning and teaching through student feedback in engineering* (pp. 43–59). Elsevier.

Skodvin, O.-J., & Aamodt, P. O. (2001). Effektivitetsbegrepet i høyere utdanning: Hvordan er begrepet brukt i norsk høyere utdanningspolitikk? (0808–4572). NIFU.

Smith, D. E. (2001). Texts and the ontology of organizations and institutions. *Studies in Cultures, Organizations and Societies*, *7*(2), 159–198. https://doi.org/10.1080/10245280108523557

Snell, L., Tallett, S., Haist, S., Hays, R., Norcini, J., Prince, K., Rothman, A., & Rowe, R. (2000). A Review of the evaluation of clinical teaching: New perspectives and challenges. *Medical Education*, *34*(10), 862–870.

Socha, A. (2013). A hierarchical approach to students' assessments of instruction. *Assessment & Evaluation in Higher Education*, *38*(1), 94–113.

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/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

Sprague, J., & Massoni, K. (2005). Student evaluations and gendered expectations: What we can't count can hurt us. *Sex Roles*, *53*(11), 779–793.

Stalmeijer, R., Whittingham, J., de Grave, W., & Dolmans, D. (2016). Strengthening internal quality assurance processes: Facilitating student evaluation committees to contribute.

Assessment & Evaluation in Higher Education, *41*(1), 53–66.
https://doi.org/10.1080/02602938.2014.976760

Stein, S. J., Goodchild, A., Moskal, A., Terry, S., & McDonald, J. (2020). Student perceptions of student evaluations: Enabling student voice and meaningful engagement. *Assessment & Evaluation in Higher Education*, 1–15. https://doi.org/10.1080/02602938.2020.1824266

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.*

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. (2006). Governmental policy, organisational ideals and institutional adaptation in Norwegian higher education. *Studies in Higher Education*, *31*(1), 43–56. https://doi.org/10.1080/03075070500392276

Stensaker, B. (2007). Quality as fashion: Exploring the translation of a management idea into higher education. In *Quality assurance in higher education* (pp. 99–118). Springer.

Stensaker, B., & Harvey, L. (2011). *Accountability in higher education: Global perspectives on trust and power*. Routledge.

Stensaker, B., Hovdhaugen, E., & Maassen, P. (2019). The practices of quality management in Norwegian higher education. *International Journal of Educational Management*, *33*(4), 698–708.

Stensaker, B., & Maassen, P. (2001). *Strategier og metoder for kvalitetsutvikling: en drøfting av noen sentrale forslag og synspunkter hos Mjøsutvalget* (0808–4572). NIFU.

Stensaker, B., Westerheijden, D. F., & Rosa, M. J. (2007). Conclusions and Further Challenges. In D. F. Westerheijden, B. Stensaker, & M. J. Rosa (Eds.), *Quality assurance in higher education, trends in regulation, translation and transformation* (Vol. 20, pp. 247–262). Springer. https://doi.org/10.1007/978-1-4020-6012-0_10

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

Stockdill, S. H., Baizerman, M., & Compton, D. W. (2002). Toward a Definition of the ECB Process: A Conversation with the ECB Literature. *New Directions for Evaluation*, *93*, 1–25.

Stokke, S., Solberg, I. L., Grøttum, P., Lundin, K. E., Heggen, K. M., & Breivik, J. (2019). Et system for studentevaluering av forelesninger ved medisinstudiet i Oslo. *Tidsskrift for Den norske legeforening*.

Strømsø, H. (2016). Forførende entusiasme: 40 års forskning på Dr. Fox-effekten. *Uniped*, *39*(2), 118–130. https://doi.org/10.18261/issn.1893-8981-2016-02-03

Forskrift om kvalitetssikring og kvalitetsutvikling i høyere utdanning og fagskoleutdanning [Regulations concerning quality assurance and quality development in higher education and tertiary vocational education], (2018). https://lovdata.no/forskrift/2010-02-01-96/§2-1

Forskrift om tilsyn av utdanningskvaliteten i høyere utdanning [Regulations concerning supervision of the educational quality in higher education], (2017). https://lovdata.no/forskrift/2017-02-07-137

Stutchbury, K., & Fox, A. (2009). Ethics in educational research: Introducing a methodological tool for effective ethical analysis. *Cambridge Journal of Education*, *39*(4), 489–504.

Sugrue, C., Englund, T., Solbrekke, T. D., & Fossland, T. (2018). Trends in the practices of academic developers: Trajectories of higher education? *Studies in Higher Education*, *43*(12), 2336–2353.

Tangen, R. (2014). Balancing ethics and quality in educational research—the ethical matrix method. *Scandinavian Journal of Educational Research*, *58*(6), 678–694.

Tavares, O., Sin, C., Videira, P., & Amaral, A. (2017). Academics' perceptions of the impact of internal quality assurance on teaching and learning. *Assessment & Evaluation in Higher Education*, *42*(8), 1293–1305.

Theall, M., Franklin, J., Theall, M., Abrami, P. C., & Mets, L. A. (2001). Looking for bias in all the wrong places: A search for truth or a witch hunt in student ratings of instruction? *New Directions for Institutional Research*, *2001*(109), 45–56. https://doi.org/10.1002/ir.3

Tran, N. D. (2015). Reconceptualisation of approaches to teaching evaluation in higher education. *Issues in Educational Research*, *25*(1), 50–61. http://www.iier.org.au/iier25/tran.pdf

Tran, N. D., & Nguyen, T. T. (2015). Shifting the focus to student learning in teaching evaluation: A new direction for Vietnamese higher education institutions? *Interchange: A Quarterly Review of Education*, *46*(2), 113–119. https://doi.org/10.1007/s10780-015-9230-1

Trigwell, K., Caballero Rodriguez, K., & Han, F. (2012). Assessing the impact of a university teaching development programme. *Assessment & Evaluation in Higher Education*, *37*(4), 499–511. https://doi.org/10.1080/02602938.2010.547929

Trochim, W. M. K. (1998). An evaluation of Michael Scriven's "Minimalist theory: The least theory that practice requires". *American Journal of Evaluation*, *19*(2), 243–249. https://doi.org/10.1016/S1098-2140(99)80202-1

UiT. (2012). *Kvalitetssystem for utdanningsvirksomheten ved Universitetet i Tromsø* [Quality assurance system UiT]. Universitetet i Tromsø. https://uit.no/om/enhet/artikkel?p_document_id=167965&p_dimension_id=88126&men=65815

UiTø. (2010). *Prosedyre for evaluering og system for eksterne evalueringer*. Universitetet i Tromsø, Det helsevitenskapelige fakultet.

Lov om Universiteter og Høyskoler [Act Relating to Universities and University Colleges], §1-6 Kvalitetssikring (2005). https://lovdata.no/lov/2005-04-01-15/§1-6

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

Varga-Atkins, T., McIsaac, J., & Willis, I. (2017). Focus group meets nominal group technique: An effective combination for student evaluation? *Innovations in Education and Teaching International*, *54*(4), 289–300. https://doi.org/10.1080/14703297.2015.1058721

Vargas-Madriz, L. F., Nocente, N., Best-Bertwistle, R., & Forgie, S. (2019). "Somebody has to teach the 'Broccoli' course": Administrators navigating Student Evaluations of Teaching (SET). *Canadian Journal of Higher Education*, *49*(1), 85–103. https://search.ebscohost.com/login.aspx?direct=true&db=ehh&AN=136553344&site=ehost-live

Varhaug, J. (2012). Studentevaluering: Hva mener studentene, og i hvilken grad bør man ta hensyn til det? In T. Løkengård Hoel, B. Hanssen, & D. Husebø (Eds.), *Utdanningskvalitet og undervisningskvalitet under press? Spenninger i høgere utdanning* (pp. s. 273–288). Tapir akademisk forlag.

Varpio, L., Ajjawi, R., Monrouxe, L. V., O' Brien, B. C., & Rees, C. E. (2017). Shedding the cobra effect: Problematising thematic emergence, triangulation, saturation and member checking. *Medical Education*, *51*(1), 40–50. https://doi.org/10.1111/medu.13124

Vasey, C., & Carroll, L. (2016). How do we evaluate teaching. *Academe*, *102*(3), 34–39.

Veeck, A., O'Reilly, K., MacMillan, A., & Yu, H. (2016). The use of collaborative midterm student evaluations to provide actionable results. *Journal of Marketing Education*, *38*(3), 157–169. https://doi.org/10.1177/0273475315619652

Vo, A. T., & Alkin, M. C. (2017). *Evaluation essentials: From A to Z*. The Guilford Press.

Von Müllen, R. (2006). Undervisningsevaluering og udvikling af underviserkompetencer. *Dansk Universitetspædagogisk Tidsskrift*, *1*(2), 16–21.

Wæraas, A., & Nielsen, J. A. (2016). Translation theory 'translated': Three perspectives on translation in organizational research. *International Journal of Management Reviews*, *18*(3), 236–270. https://doi.org/10.1111/ijmr.12092

Watson, S. (2003). Closing the feedback loop: Ensuring effective action from student feedback. *Tertiary Education and Management*, *9*(2), 145–157. https://doi.org/10.1080/13583883.2003.9967099

Weinberg, B. A., Hashimoto, M., & Fleisher, B. M. (2009). Evaluating teaching in higher education. *The Journal of Economic Education*, *40*(3), 227–261.

Weiss, C. H. (1998). Have we learned anything new about the use of evaluation? *American Journal of Evaluation*, *19*(1), 21–33.

Werr, A., & Walgenbach, P. (2019). Management techniques. In *The Oxford Handbook of Management Ideas* (pp. 104–120). Oxford University Press.

Westerheijden, D. F., Hulpiau, V., & Waeytens, K. (2007). From design and implementation to impact of quality assurance: An overview of some studies into what impacts improvement. *Tertiary Education and Management*, *13*(4), 295–312.

Westerheijden, D. F., & Kohoutek, J. (2014). Implementation and translation: From European Standards and Guidelines for quality assurance to education quality work in higher education institutions. In H. Eggins (Ed.), *Drivers and barriers to achieving quality in higher education* (pp. 1–11). Sense Publishers. https://doi.org/10.1007/978-94-6209-494-9_1

Westerheijden, D. F., Stensaker, B., & Rosa, M. J. (2007). *Quality assurance in higher education: Trends in regulation, translation and transformation* (Vol. 20). Springer Science & Business Media.

Wiers-Jenssen, J., Stensaker, B., & Grøgaard, J. B. (2002). Student satisfaction: Towards an empirical deconstruction of the concept. *Quality in Higher Education*, *8*(2), 183–195.

Wiley, C. (2019). Standardised module evaluation surveys in UK Higher Education: Establishing students' perspectives. *Studies in Educational Evaluation*, *61*, 55–65. https://doi.org/10.1016/j.stueduc.2019.02.004

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

Winchester, M. K., & Winchester, T. M. (2012). If you build it will they come? Exploring the student perspective of weekly student evaluations of teaching. *Assessment & Evaluation in Higher Education*, *37*(6), 671–682. https://doi.org/10.1080/02602938.2011.563278

Winchester, T. M., & Winchester, M. K. (2014). A longitudinal investigation of the impact of faculty reflective practices on students' evaluations of teaching. *British Journal of Educational Technology*, *45*(1), 112–124. https://doi.org/10.1111/bjet.12019

Wittek, L., & Kvernbekk, T. (2011). On the problems of asking for a definition of quality in education. *Scandinavian Journal of Educational Research*, *55*(6), 671–684.

Wolbring, T., & Riordan, P. (2016). How beauty works. Theoretical mechanisms and two empirical applications on students' evaluation of teaching. *Social Science Research*, *57*, 253–272.

Wong, W. Y., & Moni, K. (2014). Teachers' perceptions of and responses to student evaluation of teaching: Purposes and uses in clinical education. *Assessment & Evaluation in Higher Education*, *39*(4), 397–411. https://doi.org/10.1080/02602938.2013.844222

Wright, S. L., & Jenkins-Guarnieri, M. A. (2012). Student evaluations of teaching: Combining the meta-analyses and demonstrating further evidence for effective use. *Assessment & Evaluation in Higher Education*, *37*(6), 683–699.

Yin, R. K. (2018). *Case study research and applications: Design and methods* (6th ed.). SAGE Publications.

Young, K., Joines, J., Standish, T., & Gallagher, V. (2019). Student evaluations of teaching: The impact of faculty procedures on response rates. *Assessment & Evaluation in Higher Education*, *44*(1), 37–49. https://doi.org/10.1080/02602938.2018.1467878

Youssef, L. S. (2017). What can regular and timely student feedback tell us about the teaching and learning processes? *Reflective Practice*, *18*(6), 750–771. https://doi.org/10.1080/14623943.2017.1351350

Zabaleta, F. (2007). The use and misuse of student evaluations of teaching. *Teaching in Higher Education*, *12*(1), 55–76.

Zhao, J., & Gallant, D. J. (2012). Student evaluation of instruction in higher education: Exploring issues of validity and reliability. *Assessment & Evaluation in Higher Education*, *37*(2), 227–235.

Papers 1-3

Paper 1

Borch, I., Sandvoll, R., & Risør, T. (2020). Discrepancies in Purposes of Student Course Evaluations: What Does It Mean to Be "Satisfied"? *Educational Assessment, Evaluation and Accountability*, *32*(1), 83-102. https://doi.org/10.1007/s11092-020-09315-x

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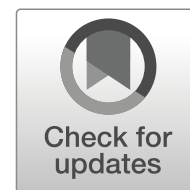
Discrepancies in purposes of student course evaluations: what does it mean to be "satisfied"?

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.

✉ Iris Borch
iris.h.borch@uit.no

Extended author information available on the last page of the article

Springer

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.ak/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) Oslu: 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
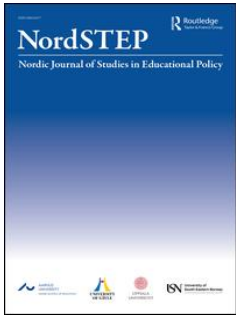
[2] The Norwegian Centre for E-Health, UiT The Arctic University of Norway, Tromso, Norway

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

Paper 2

Borch, I. (2020). Lost in Translation: From the University's Quality Assurance System to Student Evaluation Practice. *Nordic Journal of Studies in Educational Policy*, *6*(3), 231–244. https://doi.org/10.1080/20020317.2020.1818447

Lost in translation: from the university's quality assurance system to student evaluation practice

Iris Borch

Published online: 07 Sep 2020.

Submit your article to this journal ⤤

Article views: 81

View related articles ⤤

View Crossmark data ⤤

Routledge
Taylor & Francis Group

Lost in translation: from the university's quality assurance system to student evaluation practice

Iris H. Borch

Centre for Teaching, Learning and Technology, UiT the Arctic University of Norway, Tromsø, Norway

ABSTRACT

Student course evaluation is a mandatory part of quality assurance systems in Norwegian higher education, aiming to enhance educational quality. However, several studies report that student course evaluation mainly is used for quality assurance and not for quality enhancement. Drawing upon translation theory, this paper analyses how the quality assurance system (QAS) that regulates evaluation, the actors and the arenas of translation at a Norwegian university affect student evaluation practice and its uses. Academic leaders were interviewed and evaluation documents analysed. Results show that the leaders were not familiar with the university's established guidelines for an ideal evaluation practice in QAS. The academics described an evaluation practice that seems to be more internal-driven rooted in their values, previous experiences, local cultures and traditions rather than on regulations like QAS. Their translation of evaluation can be regarded as modified translation. The academics' approach to evaluation seems to be based upon a logic of appropriateness. The different actors involved in evaluation processes seem to base their actions on contradicting logics. This can help understand why a de-coupling from evaluation described in QAS occurred. These findings and the academics' perspectives should be taken into consideration when future evaluation systems are created.

Introduction

Student course evaluation has become a central part of quality assurance systems in higher education world-wide. When student evaluation was introduced in higher education in the 1960s, the aim was to use the evaluation data for improvement of teaching and students' learning (Darwin, 2016). However, rather than being a tool for quality enhancement and improved teaching and learning, these evaluations are mostly used for quality assurance of education (Darwin, 2016; Douglas & Douglas, 2006; Haji et al., 2013). Although we know that student evaluation is not always actively used to improve educational programs and students' learning (Beran et al., 2005; Kember et al., 2002; Stein et al., 2013), we lack knowledge about why this is the case. Policy makers, university management and academics consider student evaluation as an important indicator for educational quality. Consequently, the demands towards students to provide feedback about academic courses and programmes have increased (Darwin, 2016; Little & Williams, 2010) as evaluation has been incorporated in educational policies and manifested its position in regulations (Saunders, 2011). Not only have evaluation activities increased in numbers, but it also seems to be an expectation that evaluations will lead to educational quality improvement (Bamber & Anderson, 2012). This trust in evaluation might relate to the fact that evaluation is inherently rationalist and causal grounded in the logics of cause and effect (Vedung, 2010).

Although evaluation has existed within higher education since the 1960s, the formats have changed, particularly the last three decades. Whereas evaluation earlier mostly was self-regulative practices driven by the academic teachers themselves, it is nowadays often based on externally derived requirements (Trowler, 2011). This change might be explained by the introduction of management models in higher education, wherein evaluation also can be understood as a management technique, influenced by managerialism (Cuthbert, 2011). Despite this shift in regulation of educational evaluations, the actors who are responsible to conduct internal student evaluation remain the same, namely the teachers or academic leaders on programme level. In this study, their role in translation of evaluation is explored.

It is recognized that evaluation is dependent on organizational contexts (Højlund, 2014). Two organizational aspects among others that we can assume have relevance on how evaluation is organized and practised are the regulations that mandates evaluation practice

and the formal local evaluation systems. These comprise recommendations and guidelines for evaluation to direct evaluation practice.

While earlier literature reviews on evaluation use did not focus on institutional aspects (Johnson et al., 2009), evaluation approaches from this millennium such as Evaluation Capacity Building (ECB) recognize that contextual aspects in the organization play an important role in the ability to do and use evaluation (Bourgeois & Bradley Cousins, 2013; Bradley et al., 2014; Preskill & Boyle, 2008). Examples of organizational aspects contributing to increased capacity to do and use evaluations include external accountability requirements and organizational systems and structures that mediate staff interaction and communication about evaluation processes (Bourgeois & Bradley Cousins, 2013). However, there is little published research on how institutions put the policies around evaluation into play and how implementation of formal evaluation systems might affect engagement with evaluation data (Moskal et al., 2016). More knowledge about what roles the actors or stakeholders play in evaluation processes and in translating evaluation might help us understand why evaluation primarily is used as quality assurance, and not much for quality enhancement. In addition, when considering all the time spent on evaluation, research on evaluation use from the involved stakeholders' perspectives is necessary.

In this paper, I draw upon translation theory, and will investigate: How is student evaluation contextualized and translated locally at the university? More specifically, this paper analyses characteristics of (1) the QAS, (2) the arenas where evaluation takes place and (3) the actors (leaders) who are central to the planning and translation of student evaluation. Student evaluation refers to evaluations developed and initiated locally and to students' feedback about academic courses.

At this Norwegian university, actors involved in the evaluation practice are leaders, administrative staff and students, also named as key stakeholders. In this study academic leaders at health professional education programmes are interviewed. Moreover, educational documents describing student evaluation and QAS are included and analysed. The term 'arena' comprises the places where QAS and the evaluation system are established and conceptualized, hence where the idea travels *from* and the arena where evaluation and QAS travels *to*.

Regulation and use of student evaluation data in higher education

In Europe, educational evaluations in higher education institutions are frequently regulated by local quality assurance systems that comply with the European Standards and Guidelines for Quality Assurance in the European Higher Education Area (ESG) (EHEA, 2015). Within these standards and guidelines student evaluation is one of many components.

Westerheijden and Kohoutek (2014) emphasize that local implementation of the ESG should not be underestimated; cultures, norms and values in different countries and institutions are important when ideas are put into practice. They underline that various actors might understand educational reforms and management ideas in higher education in different ways. Studies of how academics regard these institutional evaluation systems state that academics accept the requirement of following these systems (Beran & Rokosh, 2009; Ory et al., 2001; Stein et al., 2012). However, other studies show that the will and motivation to use evaluation data decreases if academics believe that quality assurance systems exist to control and audit (Harvey, 2002; Newton, 2000).

The majority of research on the use of student evaluation data has been related to the validity and reliability of written evaluation methods (Hornstein, 2017; Spooren et al., 2013; Wright & Jenkins-Guarnieri, 2012) and investigations of use from teachers' perspectives (Bamber & Anderson, 2012; Beran & Rokosh, 2009; Burden, 2008; Edström, 2008; Hendry et al., 2007; Stein et al., 2012, 2013). Some empirical studies have provided insights into factors that hinder and facilitate use (Cousins, 2003; Edström, 2008; Hendry et al., 2007; Kember et al., 2002; Richardson, 2005). Although some of these factors can be categorized as organizational aspects, few have investigated their use in relation to the internal evaluation systems in higher education.

Many academics are not engaged with evaluation beyond the programme level (Edström, 2008; Hendry et al., 2007). According to recommendations from research on evaluation use, involvement throughout the whole evaluation process and a sense of ownership of the system are considered as important to enhancing the use of evaluation (Johnson et al., 2009; Patton, 2008). Therefore, a low engagement beyond programme level might affect the use of evaluation data negatively. Moreover, a study from New Zealand concludes that it is possible to increase academics' engagement in evaluation by improving technical aspects of the evaluation system (Moskal et al., 2016) and also being clear about the institutional expectations (Stein et al., 2012). In order to facilitate the use of students' feedback through staff engagement, it is recommended that universities provide

teachers consultations and opportunities to discuss evaluation findings (Neumann, 2000; Penny & Coe, 2004; Piccinin et al., 1999).

Educational policy and regulation of student evaluation in Norway

Student evaluation of teaching and programmes has been a statutory requirement in Norway since 2002. The Act relating to universities and university colleges states that all Norwegian universities are required to include student evaluation as a central part of their local quality assurance systems (Lovdata, 2005).

Student evaluation is described as essential to quality assurance of higher education in a national white paper, 'The Quality Reform', from 2001 (Meld. St. 27, 2000–2001), but has, as mentioned above, existed in higher education longer. The intention of implementing local quality assurance systems was to assure a continuous improvement of educational quality through systematic, documentation of the education programmes (Meld. St. 7, 2007–2008). Stensaker (2006) studied how six Norwegian higher education institutions, including the university in this study, adapted to political reforms that aimed to improve the quality of teaching and learning and concluded that "political pressure for reform can be difficult for higher education to reject, but that policies can be translated in various ways due to the different and sometimes conflicting 'organisational ideals'. Fifteen years after the reform, a new white paper, 'Quality Culture in Higher Education' (Meld. St. 16, 2016–2017) states that the quality assurance systems, including student evaluation, have not been used in quality development to the extent that the Ministry expected when they were introduced. The Norwegian Ministry of Education points to weaknesses with many of the quality assurance systems and expects a stronger emphasis on use of students' feedback in development of educational programmes in the future. There are few explanations and no analysis within the white paper as to why academics do not use student evaluations as expected in quality development (Meld. St. 16, 2016–2017).

Evaluation as institutionalized phenomenon

Student evaluation seems to be institutionalized in higher education today, meaning that it is a phenomenon that is taken for granted in the sector. Moreover, people expect evaluation to be an activity that takes place in modern organizations.

This study is grounded within institutionalism. Institutionalism can be considered as a way of thinking about social life and a result of human activities (DiMaggio & Powell, 1991). When referring to human activities, actions and decisions within an institutional framework, March and Olsen (1996, pp. 251–252) state that: 'choice' (…) is based more on a logic of appropriateness". Implementation of new practices or procedures is based on subjective interpretations by actors. These interpretations are influenced by established cultures and norms in an organization (DiMaggio & Powell, 1991) where actors do what they consider as appropriate within the organizational context and in relation to their role (Eriksson-Zetterquist, 2009).

Scott (2014, p. 57) defines institutions as 'multi-faceted, durable, social structures, made up of symbolic elements, social activities and material resources'. Højlund (2014) states that evaluation fits well with this definition of institution and refers to Dahler-Larsen, who claims that 'evaluation has become an institution in our society' (Dahler-Larsen, 2011, p. 3) and can be considered as an 'institutionalized standard'. Moreover, Dahler-Larsen (2006) emphasizes that the extent of institutionalization differs from organization to organization. Central building blocks or pillars in institutional structures are regulative, normative and cultural-cognitive elements, which are all build upon different logics. The regulative pillar is built upon a logic of instrumentality, the normative pillar upon a logic of appropriateness and the cultural-cognitive pillar upon a logic of orthodoxy (Scott, 2014). Czarniawska and Joerges (1996) argue that when ideas are travelling, they must be translated into local contexts. This study investigates how evaluation is translated within the local context of a Norwegian university. Because evaluation already has travelled into the university and is regarded as an institutionalized phenomenon, the paper analyses the contextualization and intra-organizational translation within the university.

Translation of evaluation as institutionalized phenomenon

The analytical framework is based on translation theory, an understanding of translations founded in institutionalism. Within institutionalism, translation is a generalized operation or process, more than a linguistic phenomenon (Eriksson-Zetterquist, 2009). 'Translation theory is characterized by a strong empirical orientation towards revealing, understanding and explaining what really happen to management ideas throughout the transfer and implementation processes (Røvik, 2019, p. 129)'

This paper draws upon an understanding of translations described by Røvik (2007), Røvik (2019, 2016, 2011). Røvik (2011) defines translation as 'more or less deliberate transformation of practices and/or ideas that happens when various actors try to transfer

and implement them'. Furthermore, Røvik (2016) describes knowledge transfers between source and recipients as acts of translation, wherein both organizational context (arena) and the participants (actors) in these processes are central to how translations are made. The actors are not passive receivers but active translators (Røvik, 2019). Moreover, translations are dependent on existing translation competence, wherein both human and institutional components are central. Translation competence refers to the translators' and organizations' capacity to shape ideas adopted from external sources into local contexts (Røvik, 2019, p. 131). Earlier translation research maintained a focus on how management techniques change in the process of application from one context to another, but there is rather little research on how translation competence affects translations (Werr & Walgenbach, 2019). All actors or stakeholders involved in student evaluation can be regarded as translators of evaluation. This study explores how actors translate evaluation within the university, after it has been institutionalized, particularly from the perspectives of the academic leaders.

Røvik (2016, p. 7) refers to three modes of translations and each of these modes has rules that characterize the translations. These modes are: the reproducing mode, the modifying mode and the radical mode. The modes can be understood as analytical distinctions to help understand translation processes between a source and a recipient.

In the context of student evaluation, *the reproducing mode* can be a programme that copies another programmes' survey and transfers it to their own context without changing anything. Central to the reproducing mode is adopting and reproduction. In *the modifying mode*, addition and omission are central rules of translation, in which addition refers to adding elements to the source version during the translations to the recipient, and omission to toning down elements. The object of translation can be a programme or course evaluation that is based on an existing evaluation but is adjusted or modified in the transfer to a new context or another course. The third mode, *the radical mode*, is a translation that is radically different from the source, i.e., a translation that is inspired by other practices (Røvik, 2016).

Evaluation use

Henry and Mark state that 'use is a core construct in the field of evaluation' (Henry & Mark, 2003, p. 293). Evaluation use is an essential part of evaluation theories and research, as well as a goal identified by most evaluators (Preskill & Caracelli, 1997). Michael Quinn Patton introduced Utilization-Focused (UFE) Evaluation in 1978, principles from which have been central to research about evaluation use and approaches to evaluation that aim to increase its uses (Patton, 2008, 1997). Central to UFE is that evaluations should be judged by their utility and actual use for intended users (Patton, 2008, p. 37). Moreover, it also emphasized that everything that happens in the evaluation process, from the beginning to the end, will affect use (Patton, 2008, p. 20). Alkin and Taut (2003) divide evaluation to use in two distinct aspects of use: findings use, and process use. Evaluation use was previously chiefly concerned with utilizing findings collected by different evaluation methods, also known as findings use. However, newer approaches to evaluation use also regard the learning that takes place during the evaluation process – process use – as an essential part of evaluation use (Johnson et al., 2009). Evaluation use in this study refers to use based upon descriptions made by leaders and in documents.

Methods

Eight health professional education programmes are included in the research. Leaders at programme level were interviewed by the author in semi-structured interviews. All informants received written information about the project this study is part of and its overall aim. The information letter also contained an informed consent, information about ethical approvals and that participation in the study was voluntary.

The leaders were included strategically and the inclusion criteria were: experience with teaching in academia, responsibility for a minimum of one academic course and experience with designing, distributing and/or summarizing student evaluations. Two of the leaders were programme leaders and consequently more involved in programme evaluation than the other informants who were responsible for only one or more courses. In this paper, the leaders are referred to as academic leaders, programme leaders or simply leaders, despite their different positions at the university. The interviews lasted 75–90 minutes and were based upon an interview guide that consisted of topics like regulations and origin of evaluation, their role in the different stages in the evaluation processes and uses of evaluation.

Educational documents from 2013 to 2015 describing evaluation practice and the system that regulates evaluation were included and analysed. These documents were from different university levels: programme, departmental, faculty and top level. From programme level, the study included evaluation templates, evaluation reports and educational quality reports. Documents included from departmental, faculty and level one, were annual educational reports documenting educational quality of the total educational portfolio at each level of the

organization. Additionally, the study included meeting agendas and board minutes from programme committee and/or departmental meetings where the evaluation and educational reports were presented and discussed. Moreover, were relevant documents concerning QAS from the university board, such as meeting agendas, board minutes, information letters about approvals and renewals of QAS to the faculties included. The documents were collected with help from administrative staff. As the author is employed at the university of the study, this probably affected the access to the documents positively (Mercer, 2007).

Leaders of programmes, departments and faculty were informed about the project early in the project period. Ethical approval was granted from the university and The Norwegian Centre for Research Data (NSD). All informants signed a consent form, the leaders and the programmes are anonymized in the presentation of the data by letter identifications A-H.

Analysis

Interview data were audio recorded and transcribed verbatim. The author did an inductive-abductive thematic analysis of the interview data in different stages in NVivo. The thematic analysis was an iterative process, but with three main stages. Each interview was analysed one by one in the first stage. In this stage, the analysis was inductive and the empirical data were sorted by codes that described the data. Descriptive and process coding were the dominant code types (Saldaña, 2013). Examples of descriptive codes were evaluation responsibility and lack of time and examples of process codes were development of evaluation tools and evaluation follow-ups. This type of codes was used to create an overview of the evaluation practice and illustrate its characteristics for the eight programmes.

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. These categories were less descriptive than those in the first stage and the thematic analysis was more abductive because the coding process was also informed by theory, like foundations for evaluation practice, feedback expectations and organizational structures. By using the categories created in the second stage, themes were developed in the last stage of the process. Throughout the process, interview data and evaluation documents from the same programme were compared to create a broader picture of the evaluation practice for each programme. Although three main stages described this process, the stages overlapped in an iterative process.

Throughout the research process, the evaluation documents played different roles and were analysed and used differently. This is expedient because the documents serve a variety of purposes (Bowen, 2009). Before each interview, templates of evaluations were read by the researcher in order to provide contextual background information about the evaluation practice at each programme. Moreover, after the interviews were conducted, documents that could provide insight and knowledge about student evaluation that the informants did not have were included and analysed as supplementary data. These documents were particularly related to documentation and use of student evaluation data on higher levels in the organization, and, furthermore, to information about how and by whom QAS was developed, formally approved and communicated at the university. This directionality in the data collection, when research questions lead to the relevant documents is considered a pragmatic approach common in projects with a constructivist orientation (Justesen & Mik-Meyer, 2010).

This study builds upon Atkinson's and Coffey's perspective about documents' role in an organization (Silverman, 2011); they view documents as artefacts that actively construct the organization they purport to describe. Moreover, they say: 'analysis therefore needs to focus on how organizational realities are (re)produced through textual conversations' (Silverman, 2011, p. 77). Cooren (2004) emphasizes that researchers often overlook that documents and texts also do something with the organization they are part of. He calls this textual agency. In this study, it is not purely the linguistics in the texts that are relevant, but foremost the textual agency, i.e., how evaluation is interpreted and documented.

Results

Before presenting the results, a short contextual overview of the university and the organizational leader structure is provided. The Arctic University of Norway has about 16 000 students at the graduate and undergraduate level and 3 600 employees organized in eight faculties. The university is structured with a certain hierarchy: a university management consisting of a rector team and a university director on top, followed by those at the faculty, department, programme and course levels. The faculty and department levels have both administrative and academic leaders, whereas the leaders of programmes are academic leaders. The leaders interviewed in this study are academic leaders at programme and/or course level where evaluation takes place.

The empirical data are presented in the following categories: *The evaluation system* and *Translation of evaluation*. Moreover, translation of evaluation is divided into: Sources for translation, Little

communication about evaluation and Need for knowledge and support. The first two subcategories refer to *arenas* for translation. The last three subcategories refer to aspects of the *actors* or the *translators*, who in this paper are actors involved in student evaluation, particularly leaders at programme level.

The evaluation system

Student evaluation is regulated by the local QAS at the university. The prevailing QAS when this study was conducted was established in 2009, approved by the Norwegian Agency for Quality Assurance in Education (NOKUT) in 2012 and revised in 2011, 2013 and 2015. When QAS is referred to in this paper, it is the 2012 version which was the existing one during the time span of this study. The revisions after 2012 were minor and not affecting the text regarding student evaluation. The development and renewal of QAS was led by designated administrative staff members. The final version and the renewed versions were approved by the university board.

As soon as a renewed version of QAS was approved by the university board, it was communicated in letter format on behalf of the university management to the faculties and published on the university's webpage. There was no information in the letter as to what the university management expected the faculties to do with the information in 2009, 2011 and 2013. The letter from 2015 included a call to the faculties about reading the details in the revision closely, and making sure employees at faculty and department level received the information about the renewed QAS.

The objectives of student evaluation are described as follows in QAS:

> Internal evaluations contribute to giving the students an active role in the work concerning the quality of education, leads to a greater focus on the student's total learning environment and to entrenching efforts concerning the quality of education in the academic environments. Evaluation is part of the students' learning process and the academic environments' self-evaluation. (Universitetet i Tromsø, 2012)

QAS allows the programme management or course leader at each unit to choose a suitable evaluation method with pertinent evaluation questions. It is possible to choose between or combine written and dialogue-based evaluation methods. The QAS encourages educators to select an evaluation method that ensures stakeholder involvement and good processing of the data material.

The frequency and timing of conducting different types of student evaluations are regulated in QAS:

> All courses must be evaluated a minimum of once every third year. (…) As a normal rule, continuous evaluation is recommended. (…) Student evaluation of courses shall be conducted during the teaching semester. (…) An annual evaluation of the programme of study shall be undertaken (Universitetet i Tromsø, 2012).

The department management is responsible for conducting evaluation of courses and following up on evaluation results, but can delegate this responsibility to the programme management. Furthermore, the programme management is responsible for conducting and following up on programme evaluations.

QAS contains guidelines about implementation and documentation, and it states that the evaluation results shall be documented and available for the students, though it does not state how. Further, it is stated that the university must have routines for analysis of the findings and provide comment on these before they make the results available for the students. Moreover, QAS states that programmes shall establish routines for how to follow up on the evaluation results. Annual reports describing educational quality including evaluation shall be written at programme, departmental and faculty level (Universitetet i Tromsø, 2012).

Translation of evaluation

Sources for translation

The leaders were in the interviews asked what regulated evaluation practice and what they based their evaluation approach on. It became clear that they were not familiar with the details in the local quality assurance system or the regulation of student evaluation. When they referred to QAS it was solely that it is mandatory to conduct student evaluation regularly – three of them mentioned the required minimum of every third year. Four of the leaders answered that they regretted to say they did not know the details of the quality assurance system (Informants B, D, G and F). One leader stood out because he replied that he knew the local QAS well (Informant C). However, this seemed mainly related to the frequency of evaluation, while he later in the interview revealed that he was not familiar with details in the QAS like the requirements to share evaluation results with students.

The interviews with the leaders uncovered uncertainty about who is responsible for follow-up on the evaluation results. They pointed to leaders or programme committees on higher organizational levels as responsible for implementation of the evaluation results. Unlike the others, one of the informants regarded himself as responsible for follow-up on the results at the programme level (Informant C).

The leaders said evaluation practice was based upon traditions, culture and previous experiences. Some of the leaders said evaluations had been conducted in the same way over many years, but the formats differed from programme to programme. Whereas some programmes had a tradition and culture of dialogue-based evaluation, other programmes had a tradition of using surveys. The leaders created written surveys with questions from templates provided by the administrative staff, copied each other's questionnaires, or formulated questions they believed would work for the courses. Moreover, they expressed that this was not a satisfying evaluation practice and elaborated how they had adjusted their approach to evaluation based upon experiences with poorly designed evaluation tools.

Lack of communication about evaluations

The leaders expressed how evaluation results barely were a topic in staff meetings or in discussions with their peers. Therefore, they had little or no knowledge about how their colleagues conducted evaluation or how other courses within the same programme were evaluated by the students. As an exception, one programme had meetings with student representatives and course leaders each semester in which evaluation was discussed. The course leaders did not know if an overall programme evaluation was conducted yearly. All the leaders desired a more shared evaluation practice, as opposed to today's practice, which two informants described as a 'lonely' part of the job (Informants E & H) and another as a 'private practice' (Informant D).

Leaders requested spaces to discuss evaluation results at the faculty. In programmes with no established forums to debate educational quality, ad hoc evaluation meetings are established. One leader said he once presented evaluation findings in the research group, or what he described as a mix between a research group meeting and a meeting with supervisors 'because many of the same colleagues are involved in both activities (…) we have no structure to discuss teaching and therefore we must use different forums' (Informant H). He expressed that too much responsibility was placed on each course leader in evaluation design, implementation and use, and had many times addressed a need for meetings to discuss education-related topics at the department.

A sufficient amount of time to do evaluations was a factor that leaders suggested was important in order to conduct and follow up on evaluations. Two of the leaders shared that evaluations were not a priority in busy times and requested more allocated resources in order to improve evaluation practice (Informants D and F). Both of them believed in a more systematic approach to evaluation – a system with reminders of

when to conduct them and a request to report on evaluation findings. They thought this could be helpful in order to prioritize evaluations in busy times of the year (Informant D and F).

In the interviews, the informants were asked if they shared evaluation results with their leaders or the students. Two of the leaders had annually contributed evaluation data to reports describing the educational quality of the programme (Informants C and G). The other leaders referred to unclear routines and systems for reporting evaluation results and were not familiar with how and if evaluation findings were reported to the next levels. One leader said he had colleagues who had lost motivation to conduct evaluation because they believe evaluation reports are simply archived. Moreover, he elaborated that his experienced student evaluation as more useful for educational improvement in those cases when the results were discussed with colleagues (Informant D).

The leaders said that there are no established routines to share evaluation results with the students; neither were there established plans for how students' feedback would be followed up. However, two of the programmes publish a summary of dialogue-based evaluations on Fronter (Learning Management System) (Programme A and B). Regarding transparency about implementation of the findings, one leader said: 'We have a potential to improve'. He had as a student himself at another university experienced getting feedback on an evaluation he had participated in. The response included students' feedback with comments and a plan for how the university intended to use the results. He valued the response and suggested that this kind of feedback was something to strive for when he said the evaluation practice could improve (Informant G).

Need for knowledge and support

The leaders expressed a need for more knowledge about evaluation and support throughout the evaluation process. One informant suggested including student evaluation as a topic in courses for new employees at the university (Informant F). Another leader referred to the design of student evaluation surveys and said: 'I wish I could work together with somebody that knows more about evaluation than me. Today, it feels like trial-and-error' (Informant G).

Yet another leader expressed a need for support in dialogue-based evaluations and implementation of students' feedback. He had once invited the students to a dialogue about the evaluation results after a course was poorly evaluated, but experienced challenges in the discussion. He said, 'If I am going to do it again, I would like to have somebody with more competence about evaluation or pedagogics with me'

(Informant D). Moreover, this leader pointed to the need for more communication with students and colleagues about evaluation results at the programme she represented.

Discussion

In order to get insight about how evaluation is contextualized and translated within the university, this section of the paper focuses on characteristics of QAS, the actors and arenas involved in translation.

What characterizes the internal quality assurance system?

As stated above, QAS was developed by administrative staff and expected to be used by academics. Administrative staff and academics obviously have different roles in higher education, but they also have different time available to immerse themselves in evaluation. The administrative staff are the ones who created the structures of evaluation practice, which the academics are supposed to follow. QAS is presented on the university's webpage and thereby accessible to students and staff. Moreover, the QAS is open for contextual adaptation of evaluation practice, customized to each course or programme, instead of directing use of one standardized evaluation tool. It seems that the university has an implicit understanding that there exists an evaluation competence on the programme level and that the leaders were familiar with the QAS. When the university commissioned evaluation in QAS, they included guidelines and recommendations of how to get an optimized evaluation practice and how to use student evaluation data to improve and assure educational quality. The university thereby provided QAS as a source and tool for translation of evaluation to the leaders. Røvik (2019) describes processes/cases when 'a management idea is concretized into specific rules, procedures and routines that organizational actors are expected to follow', like instrumentalization. The development of QAS can be understood as instrumentalization of evaluation and an expectation that academics will establish evaluation practices aligned with QAS and seems to be based upon a logic of consequences.

However, this study reveals that the leaders were not familiar with the details of the QAS. Consequently, each leader created their own local evaluation practice for the course(s) they were responsible for. They followed the requirements as stated in QAS and student evaluation took place accordingly, but they did not base their translations on details or guidelines in QAS. The evaluation practice was decoupled from QAS and the system they were part of. Each leader's translation was therefore crucial for how evaluation was put into action.

In order to get a better understanding of the translation of evaluation and why evaluation is contextualized the way it is, it is necessary to take a closer look at what characterizes the actors and the arenas involved in translation.

What characterizes the actors?

This university states in QAS that programme leaders are responsible for follow-up on QAS; as a consequence, these leaders are central actors in evaluation processes and therefore also as translators of QAS. However, these leaders did not regard themselves as translators of evaluation as stated in QAS. This means that the translator role is suppressed because they do not recognize themselves as translators responsible for putting QAS into action. In translation theory, translation competence is strongly related to knowledge about the idea and the contexts this idea is translated from and to (Røvik, 2007, 2013). In this study, the leaders did not consider themselves knowledgeable about evaluation as phenomenon or idea and they had little knowledge about the context the idea travelled from. Nevertheless, they were immersed in the contexts where evaluation took place. Knowledge about the context the idea travels into is, however, regarded as the most important translation competence (Røvik, 2019).

As the leaders did not consider themselves knowledgeable about evaluation, they had to base their approach to evaluation on their own interpretation of evaluation. Moreover, they said, they based their evaluation approach on culture, previous experiences and traditions within the programme. Their actions seem to be based upon a logic of appropriateness. Their evaluation approach can be considered as what Saunders (2011) described as 'individually driven evaluation', rooted in academic values and norms, rather than top-down directed evaluative practices. Student evaluation has existed in Norwegian higher education and at this university before the law regulation and the following implementation of quality assurance systems. The actors' evaluation practice might therefore be rooted in long-existing traditions. When they described how they created evaluations, they elaborated that written evaluations were often created by copying some questions from surveys used in other courses or programmes, some questions formulated by themselves and some from templates provided by the administration. The surveys were not standardized, but rather home-grown. One of the leaders used the phrase 'trial-and-error' when he described the process of designing surveys, while others painted similar pictures and told how they used template

questions or surveys from other courses as a foundation when they created their own. In other words, they added and subtracted elements from existing tools. They did not base their evaluation upon guidelines in QAS but created evaluation in a rather pragmatic way in order to ensure that evaluations took place and were contextualized to the programme. This can also be described as an example of a modified translation (Røvik, 2016), wherein the leaders toned down and added elements based on previous experiences and traditions at the programme they represented.

The actual evaluation practice at the included programmes can, as mentioned above, be understood as an example of modified translation. In practice, the leaders did subtractions from the standard described in QAS. These subtractions appeared in the design of evaluation tools, the distribution of surveys and in how evaluations are followed up and were most likely unintended because the leaders have limited knowledge about evaluation and the source for translation. Nonetheless, the leaders expressed good intentions to conduct and follow up on students' feedback, but acknowledged that the evaluation practices had – as one leader described – 'potential to improve'. This can be related to lack of time, absence of support throughout the evaluation process and unawareness of key aspects that might increase use, some of them appearing in QAS. In translation theory, there are different explanations as to why a phenomenon – often unintentionally – is modified from the original idea during the implementation process. Some of the explanations relate to what the leaders expressed in the interviews. Examples include lack of time and capacity; thus, this might hinder them as translators from immersing themselves in new practices they want to adapt (Røvik, 2007). Another explanation to why an idea unintended is modified when it is put into action is fragmented knowledge about the phenomenon the actors are responsible for implementing (Røvik, 2013), in this case the leaders requested more knowledge about evaluation. Leaders expressed a need for more support during the evaluation process, ideally from someone with evaluation expertise. In short, they communicated a need for consultative support throughout the whole evaluation process. In research on evaluation use, support in implementation and expertise about evaluation are identified as key factors for increased use (Johnson et al., 2009). This is also the case in higher education, use of student evaluation increases if academics receive support and help to analyse student evaluation results (Penny & Coe, 2004).

Many of the guidelines and recommendations in QAS were aligned with principles in Evaluation Capacity Building (ECB) and advice proposed in research on evaluation use. Central principles to enhance use are: involvement of stakeholders throughout the evaluation process, evaluator competence, transparency and communication about evaluation (Johnson et al., 2009). However, as stated above, the leaders said there was little transparency and discussion about evaluation at the university, neither were they as central stakeholders involved in the development of the evaluation system. They called for more knowledge about evaluation – in other words, evaluation competence. Nevertheless, are they experts on the context evaluation are translated into and they distinguished between evaluation practices that worked well and those that needed to improve. They already hold evaluation competence but they request forums where their evaluation experiences and evaluation findings could be shared and discussed. Two of the academics described ad hoc meetings or discussions about evaluation findings with peers as valuable for educational development. Individual evaluation competence is an important aspect in ECB and can be obtained directly through planned ECB activities, such as training, or indirectly through 'involvement of stakeholders in processes that produce evaluation knowledge' (Bourgeois & Bradley Cousins, 2013, p. 301). The spaces to discuss evaluation the leaders ask for could be seen as a way of indirectly obtaining ECB, in which the academics themselves should be the key stakeholders.

Although QAS includes guidelines about whose responsibility it is to design, operate and document evaluation, the leaders were uncertain about who were responsible for what. QAS allows for departments to delegate responsibilities for course evaluation to programme management; in turn, this has to be clearly expressed and agreed upon. In this study, this is not the case, and uncertainty and confusion around responsibilities occur. According to Meyer and Rowan (1977), delegation of responsibilities from management to professionals is also a well-known reason why decoupling from original ideas and structures takes place. As described above, a decoupling from the system happened when evaluation was translated into evaluation practice.

What characterizes the arenas where evaluation takes place?

Student evaluation is described as a rather open phenomenon with many possible approaches in QAS. It can be regarded as an abstract idea, and consequently, it is not a surprise that evaluation exists in many formats at the university. As discussed above, the leaders did not base their evaluation upon guidelines in QAS. One explanation for why guidelines are not followed might be related to the way information about the system is communicated within the organization.

Information about QAS is distributed in a vertical line as a top-down translation within the university. Once the university board had approved a renewed QAS, the university management oriented the lower level – the faculty. A top-down orientation of an idea or a system builds upon principles from a modern rationalistic implementation process, wherein the formal hierarchy is directing a vertical structure of information flow. Hierarchical translation or movement of an idea relate both to power and structure (Ottoson, 2009). Within a hierarchical translation chain, there are expectations as to how the contextualization of an idea happens. These expectations comprise a hierarchical top-down implementation within the organization. New ideas are directed with guidelines from the management. Local versions might occur, but the management sets the direction and expects the users at lower levels to carry out the idea within a given timeframe (Røvik, 2007).

When a management assumes that information follows a vertical line with receivers of information at different levels, it is expected that the information automatically is carried out in the organization and acted upon by leaders on lower levels, aligned with a logic of consequences. This was not the case for the leaders on the programme level in this study, as they were not aware of details in QAS, nor how evaluation was dealt with on higher levels in the organization (the arena where evaluation travels from within the university). Nevertheless, they are central actors in evaluation practice at the programme level (the arena that evaluation travels into). In order to meet the requirements for evaluation, they did pragmatic translations based upon traditions and culture, within a logic of appropriateness perspective. They did not consider themselves to be knowledgeable about evaluation or QAS, and therefore based their translations upon their interpretation of the idea of evaluation. In translation theory, this kind of translation can be considered an abstract translation (Røvik, 2007). The guidelines provided by QAS are backgrounded by the local interpretation of evaluation or translation. However, the leaders established local evaluation practices, and ensured that evaluation took place due to mandatory requirements and the direction set by the management, yet without following all the guidelines created at the top. The idea of evaluation was conceptualized differently at each level. This is an example of how evaluation was translated in sequences within the organization. First, when it was established as an idea by the management. Second, when evaluation guidelines were developed and formulated in QAS by designated staff. Third, at the faculties when they created local procedures and informed the departments about these. Fourth, at the programmes, the arena where the leaders and students in this study are actors in evaluation practice.

Between the top and the bottom, several translations of evaluation have been made. The further down in the organization evaluation travelled, the more distant from the origin it became.

Meyer and Rowan described a similar travelling of ideas, where they categorized the local versions each leader created as contextual, pragmatic adaptions of an abstract idea (Meyer & Rowan, 1977). Although such pragmatic adaptations or translations might diverge from the original idea, they might still be rational decisions that meet the internal needs and local contexts at the level where the ideas are carried out.

The Ministry of Education states (Meld. St. 16, 2016–2017, p. 71) that they, through analysis of annual education quality reports from higher education institutions, have an impression that the sector has struggled in establishing well-functioning local quality assurance systems of which academics feel a sense of ownership. Consequently, the Ministry wanted to increase academics' involvement in quality assurance and encourage them to use quality assurance systems and evaluations more actively in development of academic programmes. In order to achieve these desired objectives, the legal regulations for quality assurance and audit were changed in 2016. The current regulations include a demand to use QAS more actively in quality improvement (Meld. St. 16, 2016–2017). However, the Ministry did not provide an overall strategy about how the institutions can create a stronger sense of ownership of QAS among academics. The findings in this study are aligned with the Ministry's assumption about academics having low sense of ownership of QAS. Stakeholder involvement and sense of ownership will take time to establish and are not likely to happen automatically. Mandatory requirements from government are in institutional theory regarded as a strong driver for action (Scott, 2014). Regulations will therefore be expected to play an important role in quality assurance, but as it is the academics who are responsible to carry out evaluation, their translation competence should not be underestimated when policies, ideas or systems are put into action. In the case of this university, the translators had first-hand knowledge of the arena into which the evaluation was translated and carried out. Nevertheless, was evaluation not aligned with evaluation practice described in QAS. Knowledge about the idea and the arena from which the idea travels from are regarded as important components in translation competence (Røvik, 2007, 2016). As stated above, the informants themselves said they had little knowledge about evaluation and about the details in QAS, meaning the idea and the arena from which evaluation travelled from.

Although the description of the evaluation system in QAS in itself and the intentions for evaluations

among the academics are good, it does not mean that evaluation is translated according to the organization's intentions. The findings in this study show that the arenas where evaluation takes place have not established a shared understanding about evaluation, nor a sense of ownership of QAS among the different stakeholders. Neither has the university management established an arena and a culture conductive to implementation of QAS and evaluation. This has, in turn, probably affected how evaluation is translated. Røvik (2007) states that poor translations can also be caused by weaknesses in the implementation of an idea.

In order to improve evaluation practices and create arenas for good translations, a starting point might be found in Evaluation Capacity Building (ECB) approaches. Essential to ECB is that organizations aim to build evaluation capacity and sustainable evaluation practice by strengthening organizational factors. Examples of such factors are organizational structures and systems that mediate how members in the organization collaborate and communicate with each other (Bourgeois & Bradley Cousins, 2013; Preskill & Boyle, 2008). Developing ECB in an organization has no quick fix, it will take time to establish and will require effort and time from the involved actors or stakeholders. Although the academic leaders had a positive approach to student evaluation and considered student feedback as important for their teaching (Borch et al., 2020), it was not a priority in busy times. Despite little available time to immerse themselves in evaluation, the informants believed the university had a potential to improve organizational structures that could make it easier for them to follow better up on student evaluations.

Concluding remarks

This study aims to explore how student evaluation can be carried out at a university, and how factors of the evaluation system itself, the actors and the arenas of translation affect the translation of student evaluation. Based upon the empirical data, it became evident how characteristics of the actors and the arenas are crucial to how evaluation practice appears at a university. Evaluation has travelled from one arena to another in a vertical line within the organization. The idea originates with management who sets up an ideal evaluation practice that is planned to be implemented at the faculties, departments and programmes. Moreover, evaluation has been transformed and translated on its way between different administrative levels. It is communicated by the management but practiced by academics who seem not to be familiar with its origin. Information about guidelines and recommendations about use of evaluation

findings get lost on their way from the management to the users. The actions taken by the management who enacted QAS, the administrators who communicated the systems and the academics who translated evaluation into practice, seemed to be based upon contradicting logics. Whereas management and administrators acted upon a logic of consequences, the academic leaders based their actions upon a logic of appropriateness, their own values and available time and knowledge.

Although evaluation practice is thoroughly formulated in QAS, the study describes a discrepancy between the evaluation practice stated in QAS and the actual evaluation practice. To improve evaluation practices that in turn can be used for educational quality enhancement, organizational structures that build evaluation capacity and support academic leaders throughout the evaluation process should be strengthened. By involving academics in development of evaluation guidelines, their experiences with evaluation practices could have been incorporated. The prevailing QAS at the time of the study was open to contextual adaptations and the possibility of choosing evaluation methods suitable for a given programme. However, in order to do contextual adaption and develop evaluation approaches suitable for intended purposes, knowledge about evaluation and sufficient amount of time to follow upon evaluation guidelines are necessary. If not, the idea remains rather abstract for the translators and they do, as is the case at this university, perform deliberate transformation of evaluation. As QAS was communicated and distributed to all faculties in letter format, without clear messages about what to do with the information gathered, the arena from which the idea travelled did not prepare the arena it travelled to. This can be regarded as central to how evaluation is translated. The university seems to take for granted that the intended users were familiar with the guidelines about intended use, how to conduct and follow up evaluations in QAS. As this was not the case, modified versions of evaluation are established. A de-coupling from evaluation practice described in QAS and the actual evaluation practice occurred. The university provided academics with QAS as a tool and source for translation of evaluation, but did not involve them in discussions, training or consultation. This could be an explanation to why evaluation is not used and carried out as the university intended. It seems like the policy makers and university management expected that academics were able to translate student evaluation as described in QAS without ensuring that the evaluators knew the guidelines. The translators were left to themselves in the translation process and had a pragmatic approach to evaluation. They made sure that evaluation took place and fulfilled the statutory requirements; however, evaluation practice took a different format than

the one described in QAS and in the policy documents. Each leader did their local translation of evaluation and their translation competence was essential in how evaluation was designed, implemented and operated.

These findings underline the importance of establishing evaluation capacity in the organization, as well as translation competence, if student evaluation intend to be more actively used in educational quality development in the future.

In order to get a better understanding of why student evaluation is not used more actively in educational quality development, there is a need for more research on organizational factors and evaluation capacity, including how evaluation is translated within the sector. This study has investigated translation of student evaluation, mainly from the perspectives of academic leaders at programme level; however, other leaders, administrative staff and academics also have roles in translation processes. Research on how translation is understood from other actors in translation of evaluation will add knowledge to the field.

Acknowledgments

Disclosure statement

References

Alkin, M. C., & Taut, S. M. (2003). Unbundling evaluation use. *Studies in Educational Evaluation*, *29*(1), 1–12. http://doi.org/10.1016/S0191-491X(03)90001-0

Bamber, V., & Anderson, S. (2012). Evaluating learning and teaching: Institutional needs and individual practices. *International Journal for Academic Development*, *17*(1), 5–18. http://doi.org/10.1080/1360144X.2011.586459

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. http://doi.org/10.1007/s11251-007-9045-2

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.

Borch, I., Sandvoll, R., & Torsten, R. (2020). Discrepancies in purposes of student course evaluations: What does it mean to be "satisfied"? *Educational Assessment, Evaluation and Accountability*, *32*(1), 83–102. http://doi.org/10.1007/s11092-020-09315-x

Bourgeois, I., & Bradley Cousins, J. (2013). Understanding dimensions of organizational evaluation capacity. *American Journal of Evaluation*, *34*(3), 299–319. 10.1177/1098214013477235

Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal*, *9*(2), 27–40. http://doi.org/10.3316/QRJ0902027

Bradley, C. J., 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. http://doi.org/10.1002/ev.20076

Burden, P. (2008). Does the use of end of semester evaluation forms represent teachers' views of teaching in a tertiary education context in Japan? *Teaching and Teacher Education: An International Journal of Research and Studies*, *24*(6), 1463–1475. http://doi.org/10.1016/j.tate.2007.11.012

Cooren, F. (2004). Textual agency: How texts do things in organizational settings. *Organization*, *11*(3), 373–393. http://doi.org/10.1177/1350508404041998

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). Springer.

Cuthbert, R. (2011). Failing the challenge of institutional evaluation: How and why managerialism flourishes. In M. Saunders, P. Trowler, & R. Bamber (Eds.), *Reconceptualising evaluation in higher education: The practice turn* (pp. 133–138). McGraw-Hill Education.

Czarniawska, B., & Joerges, B. (1996). Travel of ideas. In B. Czarniawska & G. Savon (Eds.), *Translating organizational change* (pp. 13–48). Walter de Gruyter.

Dahler-Larsen, P. (2006). *Evalueringskultur: Et begreb bliver til* (2 ed.). Syddansk Universitetsforlag.

Dahler-Larsen, P. (2011). *The evaluation society*. Stanford University Press.

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

DiMaggio, P. J., & Powell, W. W. (1991). Introduction. In P. J. DiMaggio & W. W. Powell (Eds.), *The new institutionalism in organizational analysis* (pp. 1–38). University of Chicago Press.

Douglas, J., & Douglas, A. (2006). Evaluating teaching quality. *Quality in Higher Education*, *12*(1), 3–13. http://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. http://doi.org/10.1080/07294360701805234

EHEA. (2015). *Standards and guidelines for quality assurance in the European higher education area (ESG)*. Brussels.

Eriksson-Zetterquist, U. (2009). *Institutionell teori: Idéer, moden, förändring*. Liber.

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. 10.1111/medu.12091

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

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. http://doi.org/10.1080/02602930600801894

Henry, G. T., & Mark, M. M. (2003). Beyond use: Understanding evaluation's influence on attitudes and

actions. *American Journal of Evaluation*, 24(3), 293–314. http://doi.org/10.1177/109821400302400302

Højlund, S. (2014). Evaluation use in the organizational context–changing focus to improve theory. *Evaluation*, 20(1), 26–43. http://doi.org/10.1177/1356389013516053

Hornstein, H. A. (2017). Student evaluations of teaching are an inadequate assessment tool for evaluating faculty performance. *Cogent Education*, 4(1), 1304016. http://doi.org/10.1080/2331186X.2017.1304016

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. http://doi.org/10.1177/1098214009341660

Justesen, L., & Mik-Meyer, N. (2010). *Kvalitative metoder i organisations- og ledelsesstudier*. Hans Reitzel.

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. http://doi.org/10.1080/0260293022000009294

Little, B., & Williams, R. (2010). Students' roles in maintaining quality and in enhancing learning: Is there a tension? *Quality in Higher Education*, 16(2), 115–127. 10.1080/13538322.2010.485740

Lovdata. (2005). *Lov om universiteter og høyskoler [Act relating to Universities and University Colleges Section 1-6]*.

March, J. G., & Olsen, J. P. (1996). Institutional perspectives on political institutions. *Governance*, 9(3), 247–264. http://doi.org/10.1111/j.1468-0491.1996.tb00242.x

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

Meld. St. 27. (2000-2001). *Kvalitetsreform av høyere utdanning, Gjør din plikt, krev din rett*. Kirke-, utdannings- og forskningsdepartementet.

Meld. St. 7. (2007-2008). *Statusrapport for Kvalitetsreformen i høgre utdanning*. Kunnskapsdepartementet.

Mercer, J. (2007). The challenges of insider research in educational institutions: Wielding a double-edged sword and resolving delicate dilemmas. *Oxford Review of Education*, 33(1), 1–17. http://doi.org/10.1080/03054980601094651

Meyer, J. W., & Rowan, B. (1977). Institutionalized organizations: Formal structure as myth and ceremony. *American Journal of Sociology*, 83(2), 340–363. 10.1086/226550

Moskal, A. C. M., Stein, S. J., & Golding, C. (2016). Can you increase teacher engagement with evaluation simply by improving the evaluation system? *Assessment & Evaluation in Higher Education*, 41(2), 1–15. http://doi.org/10.1080/02602938.2015.1007838

Neumann, R. (2000). Communicating student evaluation of teaching results: Rating interpretation guides (RIGs). *Assessment & Evaluation in Higher Education*, 25(2), 121–134. : http://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. http://doi.org/10.1080/713692740

Ory, J. C., Ryan, K., Theall, M., Abrami, P. C., & Mets, L. A. (2001). How do student ratings measure up to a new validity framework? *New Directions for Institutional Research*, 2001(109), 27–44. http://doi.org/10.1002/ir.2

Ottoson, J. M. (2009). Knowledge-for-action theories in evaluation: Knowledge utilization, diffusion, implementation, transfer, and translation. *New Directions for Evaluation*, 2009(124), 7–20. http://doi.org/10.1002/ev.310

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

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

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

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

Preskill, H., & Boyle, S. (2008). A multidisciplinary model of evaluation capacity building. *American Journal of Evaluation*, 29(4), 443–459. http://doi.org/10.1177/1098214008324182

Preskill, H., & Caracelli, V. (1997). Current and developing conceptions of use: Evaluation use TIG survey results. *Evaluation Practice*, 18(3), 209–225. http://doi.org/10.1177/109821409701800303

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

Røvik, K. A. (2007). *Trender og translasjoner: Ideer som former det 21. århundrets organisasjon*. Universitetsforlaget.

Røvik, K. A. (2011). From fashion to virus: An alternative theory of organizations' handling of management ideas. *Organization Studies*, 32(5), 631–653. http://doi.org/10.1177/0170840611405426

Røvik, K. A. (2013). Den besværlige implementeringen: Når reformideer skal løftes inn i klasserommet. In Å. Danielsen, T. Bull, & P. Arbo (Eds.), *Utdanningssamfunnet og livslang læring* (pp. 82–93). Gyldendal Norsk Forlag.

Røvik, K. A. (2016). Knowledge transfer as translation: Review and elements of an instrumental theory. *International Journal of Management Reviews*, 18(3), 290–310. http://doi.10.1111/ijmr.12097

Røvik, K.-A. (2019). Instrumental understanding of management ideas. In A. Sturdy, S. Heusinkveld, T. Reay, & D. Strang (Eds.), *The Oxford handbook of management ideas* (pp. 121–137). Oxford University Press.

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

Saunders, M. (2011). Setting the scene: The four domains of evaluative practice in higher education. In M. Saunders, P. Trowler, & R. Bamber (Eds.), *Reconceptualizing evaluation in higher education: The practice turn* (pp. 1–17). McGraw-Hill Education: ProQuest Ebook Central.

Scott, W. R. (2014). *Institutions and Organizations: Ideas, Interests, and Identities*. (4th ed.) Thousand Oaks, California: Sage.

Silverman, D. (2011). *Qualitative research: Issues of theory, method and practice* (3rd ed.). Sage.

Spooren, P., Brockx, B., & Mortelmans, D. (2013). On the validity of student evaluation of teaching. *Review of Educational Research*, 83(4), 598–642. http://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*. Ako Aotearoa.

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. http://doi.org/10.1080/02602938.2013.767876

Stensaker, B. (2006). Governmental policy, organisational ideals and institutional adaptation in Norwegian higher education. *Studies in Higher Education*, 31(1), 43–56. http://doi.org/10.1080/03075070500392276

Trowler, P. R. (2011). The higher education policy context of evaluative practices. In M. Saunders, R. Bamber, & P. Trowler (Eds.), *Reconceptualising evaluation in higher education: The practice turn* (pp. 18–31). ProQuest Ebook Central.

Universitetet i Tromsø. (2012). *Kvalitetssystem for utdanningsvirksomheten ved Universitetet i Tromsø.*

Vedung, E. (2010). Four waves of evaluation diffusion. *Evaluation*, 16(3), 263–277. 10.1177/1356389010372452

Werr, A., & Walgenbach, P. (2019). Management techniques. In A. Sturdy, S. Heusinkveld, T. Reay, and D. Strang (Eds.), *The Oxford handbook of management ideas* (pp. 104–120). Oxford University Press.

Westerheijden, D. F., & Kohoutek, J. (2014). Implementation and translation: From European standards and guidelines for quality assurance to education quality work in higher education institutions. In H. Eggins (Ed.), *Drivers and barriers to achieving quality in higher education* (pp. 1–11). Sense Publishers.

Wright, S. L., & Jenkins-Guarnieri, M. A. (2012). Student evaluations of teaching: Combining the meta-analyses and demonstrating further evidence for effective use. *Assessment & Evaluation in Higher Education*, 37(6), 683–699. http://doi.org/10.1080/02602938.2011.563279

Paper 3

Borch, I., Sandvoll, R., & Risør, T. (2021). Student Course Evaluation Documents: Constituting Evaluation Practice. *Assessment & Evaluation in Higher Education*. https://doi.org/10.1080/02602938.2021.1899130

Appendices

Approval from Norwegian Centre for Research Data

(NSD) (Appendix 1)

Approval and extension of project period from Norwegian Centre for Research Data

(NSD) (Appendix 2)

Approval to access archive data from UiT (Appendix 3)

Interview guide; students (Appendix 4)

Interview guide; academics (Appendix 5)

Information letter to informants with informed consent(Appendix 6)

Today's evaluation practice figure 2 (Appendix 7)

Systemically aligned evaluation practice figure 3 (Appendix 8)

Appendix 1-8

Appendix 1

Approval from Norwegian Centre for Research Data (NSD) (Appendix 1)

Iris Borch

Fakultetsadministrasjonen Det helsevitenskapelige fakultet UiT Norges arktiske universitet

Postboks 6050 Langnes

9037 TROMSØ

Vår dato: 12.08.2016   Vår ref: 49144 / 3 / BGH   Deres dato: Deres ref:

TILBAKEMELDING PÅ MELDING OM BEHANDLING AV PERSONOPPLYSNINGER

Vi viser til melding om behandling av personopplysninger, mottatt 01.07.2016. Meldingen gjelder prosjektet:

49144	*Evalueringers som virkemiddel i kvalitetssikring og utvikling av emner og program*
Behandlingsansvarlig	*UiT Norges arktiske universitet, ved institusjonens øverste leder*
Daglig ansvarlig	*Iris Borch*

Personvernombudet har vurdert prosjektet og finner at behandlingen av personopplysninger er meldepliktig i henhold til personopplysningsloven § 31. Behandlingen tilfredsstiller kravene i personopplysningsloven.

Personvernombudets vurdering forutsetter at prosjektet gjennomføres i tråd med opplysningene gitt i meldeskjemaet, korrespondanse med ombudet, ombudets kommentarer samt personopplysningsloven og helseregisterloven med forskrifter. Behandlingen av personopplysninger kan settes i gang.

Det gjøres oppmerksom på at det skal gis ny melding dersom behandlingen endres i forhold til de opplysninger som ligger til grunn for personvernombudets vurdering. Endringsmeldinger gis via et eget skjema, http://www.nsd.uib.no/personvern/meldeplikt/skjema.html. Det skal også gis melding etter tre år dersom prosjektet fortsatt pågår. Meldinger skal skje skriftlig til ombudet.

Personvernombudet har lagt ut opplysninger om prosjektet i en offentlig database, http://pvo.nsd.no/prosjekt.

Personvernombudet vil ved prosjektets avslutning, 30.10.2020, rette en henvendelse angående status for behandlingen av personopplysninger.

Vennlig hilsen

Kjersti Haugstvedt

Belinda Gloppen Helle

Kontaktperson: Belinda Gloppen Helle tlf: 55 58 28 74

Vedlegg: Prosjektvurdering

Dokumentet er elektronisk produsert og godkjent ved NSDs rutiner for elektronisk godkjenning.

Personvernombudet for forskning

Prosjektvurdering - Kommentar

UTVALG OG DATAINNSAMLING

Utvalget består av studenter, ansatte og ledere på ulike nivå. Datainnsamlingen består av personlige intervjuer og gjennomgang av studentevalueringer av ulike emner og program, samt rapporter som sammenfatter studentevalueringene. Studentevalueringene består av skriftlige evalueringer og referat fra dialogbaserte evalueringer. Forsker (jf. epost 11.08.16) forklarer at evalueringene i utgangspunktet skal være anonyme, men det kan ikke utelukkes at det kan fremkomme enkelte personopplysninger om studentene som har skrevet/gitt evalueringene og tredjepersoner, som ansatte, foreleser etc.

INFORMASJON OG SAMTYKKE

Den delen av utvalget som skal intervjues informeres skriftlig og samtykker til deltagelse. Informasjonsskrivet er godt utformet, men dato for prosjektslutt må samsvare med den oppgitt i meldeskjemaet. Dato for prosjektslutt må derfor endres til 30.10.2020.

Studentene som har skrevet studentevalueringene kan mest sannsynlig ikke identifiseres ut i fra den informasjonen som foreligger, ettersom evalueringene skal være anonyme. Det kan likevel fremkomme enkelte opplysninger som kan identifisere studentene og det kan fremkomme opplysninger om tredjepersoner som studentene har omtalt i sine evalueringer. Tredjepersonene kan være ansatte ved institusjonen eller forelesere.

Det er avklart med forsker (jf. epost 11.08.16) at det ved gjennomgang av evalueringene ikke skal registreres opplysninger om studentene eller tredjepersoner, verken direkte eller indirekte. Fokuset ved gjennomgang av evalueringene er ikke på personopplysninger, men på selve evalueringen. Studentene eller tredjepersoner skal ikke kunne gjenkjennes i publikasjonen. Videre vil en gjennomgang av evalueringsformene ved universitetet kunne forbedre tilbudet for det involerte utvalget. Personvernombudet finner derfor at opplysningene kan behandles i henhold til personopplysningsloven § 8 d).

ANDRE GODKJENNINGER

Personvernombudet forutsetter at universitet har tillatelse og mulighet til å dele studentevalueringene med forsker.

INFORMASJONSSIKKERHET

Personvernombudet legger til grunn at dere behandler alle data og personopplysninger i tråd med UiT Norges arktiske universitet sine retningslinjer for innsamling og videre behandling av forskningsdata og personopplysninger.

PROSJEKTSLUTT OG ANONYMISERING

I meldeskjemaet har dere informert om at forventet prosjektslutt er 30.10.2020. Ifølge meldeskjemaet skal dere da anonymisere innsamlede opplysninger. Anonymisering innebærer at dere bearbeider datamaterialet slik at

ingen enkeltpersoner kan gjenkjennes. Det gjør dere ved å slette direkte personopplysninger, slette eller omskrive indirekte personopplysninger, slette digitale lydopptak og slette/anonymisere studentevalueringene.

Appendix 2

Approval of extension of project period from Norwegian Centre for Research Data (NSD)

NSD NORSK SENTER FOR FORSKNINGSDATA

NSD sin vurdering

Prosjekttittel

49144 Evalueringer som virkemiddel i kvalitetssikring og utvikling av emner og program

Referansenummer

932485

Registrert

14.05.2020 av Iris Helene Borch - iris.h.borch@uit.no

Behandlingsansvarlig institusjon

UIT – Norges Arktiske Universitet / Universitetsbiblioteket

Prosjektansvarlig (vitenskapelig ansatt/veileder eller stipendiat)

Iris Helene Borch, iris.h.borch@uit.no, tlf: 99380404

Type prosjekt

Forskerprosjekt

Prosjektperiode

08.08.2016 - 25.09.2021

Status

29.05.2020 - Vurdert

Vurdering (1)

29.05.2020 - Vurdert

BAKGRUNN
Behandlingen av personopplysninger ble opprinnelig meldt inn til NSD 01.07.2016 (NSD sin ref: 49144) og vurdert under personopplysningsloven som var gjeldende på det tidspunktet.
Den 14.05.2020 meldte prosjektleder inn forlengelse av prosjektperioden.

VURDERING
Det er vår vurdering at behandlingen/hele prosjektet vil være i samsvar med den gjeldende personvernlovgivningen, så fremt den gjennomføres i tråd med det som er dokumentert i meldeskjemaet 29.05.2020 med vedlegg, samt i meldingsdialogen mellom innmelder og NSD. Behandlingen kan fortsette.

MELD VESENTLIGE ENDRINGER
Dersom det skjer vesentlige endringer i behandlingen av personopplysninger, kan det være nødvendig å melde dette til NSD ved å oppdatere meldeskjemaet. Før du melder inn en endring, oppfordrer vi deg til å lese om hvilke type endringer det er nødvendig å melde:

https://nsd.no/personvernombud/meld_prosjekt/meld_endringer.html
Du må vente på svar fra NSD før endringen gjennomføres.

TYPE OPPLYSNINGER OG VARIGHET
Prosjektet vil behandle alminnelige kategorier av personopplysninger frem til 25.09.2021. Opprinnelig prosjektslutt var 30.10.2020.

LOVLIG GRUNNLAG
Prosjektet har innhentet samtykke fra de registrerte til behandlingen av personopplysninger.

Vår vurdering er at prosjektet la opp til et samtykke i samsvar med kravene i art. 4 og 7, ved at det var en frivillig, spesifikk, informert og utvetydig bekreftelse som kan dokumenteres, og som den registrerte kan trekke tilbake. Samtykket vurderes som gyldig også etter gjeldende personvernregelverk.

Lovlig grunnlag for behandlingen er den registrertes samtykke, jf. personvernforordningen art. 6 nr. 1 bokstav a.

PERSONVERNPRINSIPPER
NSD vurderer at behandlingen av personopplysninger følger prinsippene i personvernforordningen om:

- lovlighet, rettferdighet og åpenhet (art. 5.1 a), ved at de registrerte har fått tilfredsstillende informasjon om og har samtykket til behandlingen
- formålsbegrensning (art. 5.1 b), ved at personopplysninger er samlet inn for spesifikke, uttrykkelig angitte og berettigede formål, og ikke viderebehandles til nye uforenlige formål
- dataminimering (art. 5.1 c), ved at det kun behandles opplysninger som er adekvate, relevante og nødvendige for formålet med prosjektet
- lagringsbegrensning (art. 5.1 e), ved at personopplysningene ikke lagres lengre enn nødvendig for å oppfylle formålet

DE REGISTRERTES RETTIGHETER
Så lenge de registrerte kan identifiseres i datamaterialet vil de ha følgende rettigheter: åpenhet (art. 12), informasjon (art. 13), innsyn (art. 15), retting (art. 16), sletting (art. 17), begrensning (art. 18), underretning (art. 19), dataportabilitet (art. 20).

NSD vurderer at informasjonen som de registrerte mottok var godt utformet under personopplysningsloven som var gjeldende på det tidspunktet. Det vurderes at informasjonen også er tilstrekkelig for å innhente et informert samtykke og oppfylle informasjonsplikten etter nytt personvernregelverk. Informasjonen som de registrerte mottok oppfyller krav til form, jf. personvernforordningen art. 12.1, og mangler kun informasjon om nye rettigheter og kontaktopplysninger til institusjonens personvernombud for å oppfylle alle krav til innhold, jf. art. 13.

Vi minner om at hvis en registrert tar kontakt om sine rettigheter, har behandlingsansvarlig institusjon plikt til å svare innen en måned.

FØLG DIN INSTITUSJONS RETNINGSLINJER
NSD legger til grunn at behandlingen oppfyller kravene i personvernforordningen om riktighet (art. 5.1 d), integritet og konfidensialitet (art. 5.1. f) og sikkerhet (art. 32).

For å forsikre dere om at kravene oppfylles, må dere følge interne retningslinjer og eventuelt rådføre dere med behandlingsansvarlig institusjon.

OPPFØLGING AV PROSJEKTET
NSD vil følge opp ved planlagt avslutning for å avklare om behandlingen av personopplysningene er avsluttet.

Lykke til videre med prosjektet!

Kontaktperson hos NSD: Eva J B Payne
Tlf. Personverntjenester: 55 58 21 17 (tast 1)

Appendix 3

Approval to access archive data from UiT

Stipendiat Iris H. Borch

Svar på søknad om tilgang til studentevalueringer ved Det helsevitenskapelige fakultet

Vi viser til brev datert 16.8.2016 der du søker om tillatelse til å innhente data knyttet til tidligere evalueringer av emner og studieprogram ved Det helsevitenskapelige fakultet. Dataene skal benyttes i ditt doktorgradsprosjekt *Evaluering som virkemiddel for utvikling og kvalitetssikring av emner og studieprogram*.

UiT tillater at du gis tilgang til de aktuelle dataene og anmoder Det helsevitenskapelige fakultet om å bistå deg i å framskaffe datamaterialet ved behov. Ta kontakt med oss igjen dersom det underveis viser seg at du trenger data som Avdeling for utdanning kan framskaffe (f.eks. fra Felles studentsystem).

Lykke til med prosjektet!

Vennlig hilsen

Julia Holte Sempler e.f.
seksjonsleder
—

Hege Svendsen
seniorrådgiver
—

hege.svendsen@uit.no
77 64 57 43

Dokumentet er elektronisk godkjent og krever ikke signatur

Kopi: Det helsevitenskapelige fakultet

Appendix 4

Interview guide; students

Thematic interview guide for focus group discussions with students

Introduction

- Presentation of focus group participants

Student's perspectives on:

1) **Student evaluation of teaching, courses and programmes**

- Purpose of student evaluation (National, locally and personally)
- Information about student evaluation from the university

2) **Student evaluation processes at the programme**

- Types of evaluation at programme (Courses, Semester, Programme, Practical placement)
    - Time of conduct (End of courses? Mid-term? Continous)
- Experiences with different types of evaluation
    - Written surveys
        - What characterizes surveys?
        - Types of questions
        - Feedback about results
        - Feedback about evaluation use
    - Dialogue-Based Evaluations
        - What characterizes dialogue-based evaluations?
        - Types of questions
        - Who are leading the discussions and how?
        - Feedback about results
        - Feedback about evaluation use

3) **Evaluation processes**

- What kind of topics do you consider as important to evaluate?
- If you designed evaluations, how would they be carried out?
- Motivation to provide feedback?
    - Total amount of evaluations?
    - Evaluation fatigue?

- How do you experience that evaluations are followed up?
  - Feedback about results
  - Feedback about evaluation use
  - Evaluation summaries provided
  - Empowerment?
  - Discussions about evaluations with peers?
  - Discussions about evaluations with teachers/educational leadership?

4) Actors in evaluation processes

- Student involvement in evaluation processes
- Evaluation as topic in meetings with teachers and educational leadership?

5) Evaluation of practical placements

- What characterise evaluations of practical placements (Format, content, student learning outcomes, supervisors)
- Which topics do you consider as important to evaluate about placements?
- How are these followed up?

6) Student evaluation and examples of evaluation uses

- Educational quality
- Quality assurance
- Quality enhancement
- Learning processes

Closure

- Any other reflections about student evaluation processes you would like to add?

- Thank you for your contribution

Appendix 5

Interview guide; academics

Thematic interview guide for semi structured interviews with academics

Introduction
- Presentation (Job title, years in academia, formal responsibility for student evaluation etc.)
- What kind of student evaluation processes are you involved in?

Academic's perspectives on:

1) **Student evaluation of teaching, courses and programmes**
- Purposes of student evaluation (National, locally, personally)
- Information about evaluation from the faculty, department and/or programme
- Evaluation competence at the programme
- What are evaluation approaches based upon (Theories, culture, tradition, philosophy)

2) **Student evaluation at the educational programme**
- Which course(s)/programme are you responsible for?
- Types of evaluation at programme (courses, semester, programme, practical placement)
   - Time of conduct (mid-term or end, each time?)
- Experiences with different types of evaluation
   - Written surveys
      - What characterizes surveys?
      - Types of questions
      - Feedback to students about results?
      - Feedback to students about use?
   - Dialogue based evaluations
      - What characterizes dialogue based evaluations?
      - Types of questions
      - Feedback to students about results?
      - Feedback to students about use?
- How are practical placements evaluated? (design, content, follow up)

3) **Evaluation processes**
- Evaluation focus/content
   - What kind of topics do you consider as important to evaluate?
   - Which topics are evaluated? (Learning outcomes, Course structure, Reading list, Satisfaction, Learning environment, Learning management system, Learning activities)
   - Strengths with today's evaluation process
   - Challenges/limitations with today's evaluation process
   - Response rates/participation

4) Actors in evaluation processes
- Who are designing/planning the evaluations? (templates)
  - If you, in collaboration with somebody? (administrative staff, programme/department/faculty leadership?)
- Describe the roles of the different actor's and stakeholder's
- Student involvement

5) Analysis, follow-up and feedback
- How do you follow up on evaluations?
  - Analysis, evaluation plans
  - Documentation of evaluation results and use
    - Procedures for sharing
    - Reports, feedback on reports
  - Discussions with colleagues
  - Evaluation follow up on department and faculty level
  - Time spent on evaluation (throughout evaluation processes from planning to reporting)
  - Motivation
  - Student involvement in analysis and follow up (direct and indirect, available evaluation reports)
  - Feedback from colleagues about evaluation practice
  - Feedback from students about evaluation practice
  - Student evaluation as topic at formal meetings (leadership meetings, administrative meetings, team meetings, meetings with student representatives)

6) Student evaluation and examples of evaluation uses
- How do you consider student evaluation affect:
  - Educational quality
  - Learning processes
- How do you consider student evaluation contribute to
  - Quality assurance
  - Quality enhancement
- Examples of evaluation use
- Examples of valuable evaluation processes
- Examples of evaluation processes that were not carried out as intended
- How would you consider student evaluation responses/data can serve as indicator for educational quality and basis for educational development? To enhance students' learning processes?

Closure

- Any other reflections about student evaluation processes you would like to add?
- Thank you for your contribution

Appendix 6

Information letter to informants with informed consent

Forespørsel om deltakelse i forskningsprosjektet "Evalueringers betydning som virkemiddel i kvalitetssikring og utvikling av emner og program"

Bakgrunn og formål

Jeg henvender meg til deg fordi du har viktige erfaringer som student/leder om evaluering av høyere utdanning. For tiden er jeg doktorgradsstudent tilknyttet UiT Norges arktiske universitet, RESULT (Ressurssenter for undervisning, læring og teknologi). Tema for forskningsprosjektet er evaluering som virkemiddel for kvalitetsutvikling og kvalitetssikring av høyere utdanning. Jeg skal forske på hvordan studentevalueringer av utdanning og emner følges opp i organisasjonen og hvilken rolle de spiller i endring/utvikling av emner og program. Per i dag har vi lite kunnskap om dette. Dine, og andre studenters/lederes, erfaringer er av stor betydning for å fremskaffe denne type kunnskap om studentevalueringers rolle i utvikling av utdanning. Jeg håper derfor du er interessert i å delta i dette forskningsprosjektet.

Hovedmålet med doktorgradsprosjektet er å fremskaffe kunnskap om hvilke evalueringsformer som anvendes ved Det helsevitenskapelige fakultet, UiT, hvordan de følges opp ved ulike nivå i organisasjonen og betydningen evalueringene har for utvikling av studieprogram/emner. Prosjektet har følgende tre forskningsspørsmål:

(I) Hvilke evalueringsformer anvendes ved Det helsevitenskapelige fakultet, UiT Norges arktiske universitet, og hva kjennetegner disse?

(II) Hvordan følges evalueringene fra programnivå opp på institutt og fakultetsnivå?

(III) Hvilken betydning har evalueringene for endringer/utvikling av utdanningsprogram/emner?

Utvalget av informanter er strategisk utvalgt, og inkluderer både studenttillitsvalgte, studenter i studentutvalg og ledere ved ulike nivå. Informantene som forespørres om å delta har alle vært deltagere på studentevalueringer eller vært med å behandle studentevalueringer.

Hva innebærer deltakelse i studien?

Dersom du takker ja til å delta i studien, vil du bli invitert til et kvalitativt forskningsintervju om evaluering av emner/studieprogram. Intervjuene med studentene vil bli gjennomført som

fokusgruppeintervju, og intervjuene med lederne som individuelle semistrukturerte intervju. Intervjuene er beregnet til å ta mellom 60-90 minutter og vil ta utgangspunkt i en tematisk intervjuguide. Det vil bli gjort lydopptak av intervjuet og intervjuer vil ta noen notater underveis.

Hva skjer med informasjonen om deg?

Alle personopplysninger vil bli behandlet konfidensielt. Etter at intervjuene er gjennomført vil lydfila overføres til passordbeskyttet pc og lydfila slettes fra diktafonen. Intervjudata vil deretter analyseres og utgjøre grunnlaget for datamaterialet i doktorgradsprosjektet. Informantenes navn vil lagres adskilt fra øvrige data. Det er kun stipendiaten og veilederne som vil ha tilgang til rådata fra intervjuene. I publisering av funn fra forskningen er alle informanter anonymisert.

Prosjektet skal etter planen avsluttes 30.10.2020. Alle personopplysninger og opptak vil slettes senest tre måneder etter kandidaten har disputert.

Frivillig deltakelse

Det er frivillig å delta i studien, og du kan når som helst trekke ditt samtykke uten å oppgi noen grunn. Dersom du trekker deg, vil alle opplysninger om deg bli slettet.

Dersom du har spørsmål til studien, ta gjerne kontakt med undertegnede. Hvis du ønsker å delta i prosjektet og kan stille på intervju, ber jeg deg svare på e-posten. Samtykkeerklæringen nederst i brevet må signeres før intervju, og returneres per e-mail, post eller når vi om mulig møtes for intervju.

Veiledere er: Førsteamanuensis Ragnhild Sandvoll (hovedveileder), ragnhild.sandvoll@uit.no
Førsteamanuensis Torsten Risør (biveileder), torsten.risor@uit.no

Studien er meldt til Personvernombudet for forskning, NSD - Norsk senter for forskningsdata AS.

Med vennlig hilsen

Iris H. Borch
Stipendiat/prosjektleder

RESULT, UiT Norges arktiske universitet

 iris.h.borch@uit.no Tlf: 77 64 45 42

Samtykke til deltakelse i studien

Jeg har mottatt informasjon om studien, og samtykker ut fra denne til å delta i studien

----

(Signatur prosjektdeltager, dato)

Appendix 7

Today's evaluation practice figure 2

Appendix 8

Systemically aligned evaluation practice figure 3

System alignment

Continuous analysis, feedback and improvements

Constructive alignment

Admin
Academics/teachers
Students

Learning and teaching activities

Academics/teachers
Students

Learning oriented assessment

Academics/teachers
Students

Admin

Student learning oriented documentation of educational quality

Curriculum with learning outcome descriptions

Quality assurance system and the guidelines for student evaluation

Academics/teachers
Students
Admin

Evaluation planning

Formative evaluation

Summative evaluation

Evaluation reporting & feedback

Student-learning-centred evaluation system

Admin
Acadmics/teachers
Students