

Becoming an Excellent Teaching Practitioner: A Study of Assessment Design in a Higher Education Institution

Vidar Gynnild and Astrid Gynnild

Abstract

In higher education worldwide, the terms ‘teaching excellence’ and ‘educational merit systems’ are frequently used in institutional policy documents. Following international assessment trends, Nordic universities and colleges have for decades encouraged educators to apply for the Excellent Teaching Practitioner (ETP) award. However, little is known about rejected applicants’ experiences with such assessment processes, and how such knowledge could contribute to further quality improvement. This thematic analysis of an ETP process at a large Nordic higher education institution shows that social–psychological effects of such processes ought to be investigated more closely. This study indicates that applicants were shocked and distressed by the discursive opaqueness and a lack of alignment in the assessment design. Administrative maneuvers to standardize the assessment design came into conflict with academically established evaluation principles such as transparency, authenticity, validity, reliability, and fairness.

Keywords: assessment design; assessment policy; educational merit systems

Received 19 August 2024; revised version received 20 March 2025; accepted 20 March 2025. Corresponding author: Vidar Gynnild, Norwegian University of Science and Technology, Norway (Vidar.Gynnild@ntnu.no).

Introduction

Although the concept ‘teaching excellence’ is frequently found in university policy documents, there exists no shared understanding of its practical interventions (Gunn & Fisk, 2013). The Excellent Teaching Practitioner (ETP) award at Nordic higher education institutions is a merit system pedagogically inspired by the American Scholarship of Teaching and Learning (SoTL) tradition (Aurvoll et al., 2023; Felten, 2013; Førlund & Roxå, 2024; Miller-Young & Yeo, 2015). This ETP title is often accompanied by a monetary award, whereas the assessment design varies considerably between institutions. When the award was introduced in the new millennium, following international neoliberal motions, governments’ stated goals were to credential excellent teaching and promote educational development

(Ministry of Education and Research, 2017).

More recently, however, the teaching excellence scheme has become a controversial topic among academics across countries; many educators are critical of what such evolving merit systems evaluate in practice. Unfortunately, research-based knowledge about the policy, processes, and impact of teaching excellence systems remains scarce. Focus is typically on successful merits; therefore, we know little about potential pitfalls of ETP assessments. In this study, we investigate the ETP assessment design at an internationally oriented Nordic research and teaching institution. The immediate reason for this study was the published prevalence of rejected applicants. By investigating experiences of applicants who failed in the process, this study identifies and analyzes hitherto unknown implications of an evolving ETP merit system. Based on the three themes that emerged from the thematic data analysis, we discuss how the apparent mismatch between academic assessment principles and neoliberal ideology can be conceived and mended.

Theoretical background

The new millennium marked the beginning of a new education-focused era in higher education at the international level. Many universities embarked on a mission to systematically raise the status of teaching and learning. In the first phase, the new approaches focused primarily on innovative learning activities but did not address the complex issues concerning educational merit (Chalmers, 2011; Hutchings & Shulman, 1999). Rather, the pedagogical literature around the turn of the century suggested that, for teachers in higher education, the best step forward was to initiate educational development projects for the benefit of students (Casey et al., 1997).

Constructive alignment (Biggs, 1996) drew on constructivist learning theory; the basic idea was to align learning design components with desired outcomes. Constructive alignment quickly became a key concept in Western course and study program designs, serving as a pedagogical lodestar for policy reform and educational development over the next few decades. Teaching design was transformed from curricula-based to outcome-based routines, which have been the hallmark of teaching and learning activities in higher education ever since.

Although academic institutions wanted to pinpoint their engagement in teaching and learning, the criteria for assessing best teaching practices remained unclear (Subbaye & Vithal, 2017, p. 37). A wide range of ‘excellence in teaching’ initiatives were introduced to rebalance what several teaching-oriented academics, especially in college-level higher education, perceived as the dominance of research (Gunn & Fisk, 2013; Subbaye & Vithal, 2017). A variety of award schemes were established to promote a culture of quality in higher education while simultaneously

demonstrating commitment to activities important to student learning (Ka Yuk Chan & Chen, 2024).

In 2016, it was stated that the United Kingdom would ‘introduce a Teaching Excellence Framework (TEF) to provide clear information to students about where the best provision can be found and to raise the standard of teaching in all universities’ (Wood & Su, 2017, p. 451). In the United Kingdom, the TEF acts as a national framework for assessing undergraduate teaching, and countries on other continents have developed similar programs such as the Awards for Teaching in Australia and the University Grants Teaching Award in Hong Kong (Ka Yuk Chan & Chen, 2024).

While commitment to excellence seems to be a priority within higher education, critics argue that the quest for teaching excellence is unjustifiably presented as an a priori ideal in the higher education sector: ‘Yet, excellence is nothing in and of itself; it is only the measurement of something else’ (Saunders & Blanco Ramírez, 2017, p. 396). Evidence of ‘excellence’ requires an instrument that breaks down complex intellectual issues into discrete and observable units. Therefore, the measurement of a construct such as teaching excellence is extremely demanding, and those involved run a constant risk of oversimplifying the multifaceted nature of ‘teaching’ (Saunders & Blanco Ramírez, 2017, p. 397).

Skelton (2004) pointed out that definitions of excellence are diverse and context-dependent and that it would not be possible to agree on a single definition of what it means to be an excellent teacher. To exemplify his argument, he proposed four different approaches to excellence in teaching: The traditional approach focused on the mastery of the discipline; the performative approach aligned with the notion of meeting standards and quality thresholds; the psychological approach focused on teacher–student interrelations and could result in lifelong learners; and the critical approach concerned commitment to social justice. Furthermore, according to Wilcox (2020), ‘competencies, orientations and priorities can differ across contexts and disciplines, often contradicting generic models of excellent teaching’ (p. 45).

The application of employee recognition programs in general is highly criticized due to their roots ‘in the profound transformation of higher education caused by processes of neoliberalism, driven by economic imperatives to develop “global, entrepreneurial, corporate, commercialized universities”’ (Gourlay & Stevenson, 2017, p. 391). Hancock (2022) pointed out that reward and recognition schemes function to reinforce desired behaviors and motivate academics to work harder. Others have claimed that the neoliberal university takes competition ‘for granted, and therefore metrics for comparison are evermore necessary’ (Saunders & Blanco Ramírez, 2017, p. 396). The subsequent focus on measurability is criticized for substituting metrics for ‘excellence’ (Wood & Su, 2017, p. 452), and it is argued that the measurement approach is reductionist as it ignores

organizational complexities. It is an ‘assault on the essential dignity and autonomy of the individual’ (Hancock, 2022, pp. 381–382).

In a comparative study on assessment policy discourses, Raaper (2017) found that, at Tallinn University in Lithuania, a key assessment principle—accountability—had academics as its assessors. Upon searching for other generic principles, such as transparency, consistency, and fairness, she found Tallinn University to be run in accordance with a collegial governance model. In contrast, at the University of Glasgow, the assessment policy drew more on ambiguous language use, standardization, and diffuseness. More agents, human and nonhuman, were involved in the assessment processes, and assessment as a disciplinary power had become more diffuse.

A Nordic perspective

In the Nordic countries, the idea of educational merit systems, understood as measures of teaching excellence, was first adopted by Sweden around the turn of the century (Lindberg-Sand & Olsson, 2008; Mårtensson et al., 2011), and spread from Sweden to Norway and Finland (Aurvoll et al., 2023; Raaheim et al., 2020; Sandvoll et al., 2018). By 2016, more than half of Swedish higher education institutions had established merit systems based on a common institutional model (Winka, 2017). Different institutions, and sometimes even faculties within an institution, employed different merit models, although they all applied the title ‘Excellent Teaching Practitioner’ (Mårtensson et al., 2011; Olsson et al., 2010; Olsson & Roxå, 2013; Winka, 2017). Norwegian policy reforms included requirements for ‘educational merit systems’ to credential excellent teaching and stimulate further educational development (Ministry of Education and Research, 2017).

Many universities announced that university teachers could apply for the ETP distinction every year, or every other year. However, in the investigated case, the next application round was postponed for several years due to internal criticism of the institution’s teaching excellence practices. In a report (Raaheim et al., 2020), it was concluded that there was still extensive work for the institution to do in terms of clarifying the intended outcomes of the ETP scheme and its actual content. Applicants perceived the criteria as ‘unclear,’ and feedback to the applicants was considered short and insufficiently informative.

Based on these findings, the ETP scheme’s evaluation committee suggested a clearer structure in the form of a ‘recipe’ for the application process (Raaheim et al., 2020, p. 17); however, no reference to international research was made to support this recommendation. The revision of the merit resulted in, among other things, a new set of criteria and webinars to support the application process.

Assessors were recruited both internally and externally and firmly instructed how to assess the applications.

The revision of the ETP merit project was a top-down initiative, as it was initiated and led by administrators. It was developed in cooperation with consultants who enjoyed a long history in running ETP systems in higher education. These consultants also assisted in establishing the assessment design on which successful applicants were granted a monetary award of 5000 Euros.

Limited research

Research-based knowledge about Nordic merit systems of teaching excellence is scarce. Academics and governments have repeatedly emphasized the need for more systematic studies on the criteria for teaching excellence in higher education (Aurvoll et al., 2023; Larsen Damsgaard et al., 2024; Sandvoll et al., 2018). A recent study by Larsen Damsgaard et al. (2024) found that a main concern of academics involved the unwanted consequences of new public management and neoliberalism, such as expanding bureaucracy, top-down control management, standardization and procedures, more rules, and less academic freedom. While educators increasingly perceived the management of higher education institutions to be nonacademic, they felt overwhelmed by time-consuming administrative tasks at the same time.

Internationally, little is known about the assessment design and the utility of excellence awards. Endeavors are perceived to have a positive impact, but few studies exist to support such claims; furthermore, information on unintended consequences is scarce (Huggett et al., 2012). In a time of neoliberalism and rising competition for students, leading international universities have invested time and effort in developing assessment policies. Such policies are considered a means to create positive cultures and a shared understanding of principles associated with sound assessment practices. These policies are naturally of interest to students and their choice of educational institution.

A comparative study of the assessment policies of leading universities on four continents, involving the countries Australia, South Africa, Hong Kong, and the United Kingdom, found assessment policies to be focused on how principles are organized into a coherent system to support transparent, valid, reliable, authentic, and fair decisions (Gynnild, 2017). This study confirmed the adoption of a shared set of generic assessment principles serving as guidelines at all levels of the examined institutions. Such assessment policies are typically presented digitally, for instance as seen at the University of Cape Town, South Africa (University of Cape Town, 2024).

Research questions

As the total number of applicants for the ETP scheme are typically very low, often less than 50 at institutions with thousands of university teachers, our hypothesis was that those who actually applied would be particularly passionate about teaching and learning and that most of them would qualify for the ETP distinction. The rejected applicants at the institution in this study thus represented an interesting target group, and we were curious about how they had experienced the entire application journey. What were their experiences with the assessment process, and what lessons could be learned from these experiences? In this way, we also wanted to contribute with more knowledge about how an academically sound and well-functioning ETP assessment process suitable for educators in higher education should or could be designed.

Materials and methods

This analysis is based on qualitative data from oral and written sources collected at a Nordic higher education institution hosting more than 20,000 students and several thousand educators spread across different campuses. Recently, more than half of the ETP applicants failed, a notable affair which motivated this study of rejected applicants' experiences with the assessment process. As an exploratory study, it was carried out as a stepwise, inductive endeavor. First, we familiarized ourselves with national and international literature on teaching excellence and educational merit schemes. Second, we conducted five qualitative, open-ended interviews with rejected applicants at the institution under scrutiny. A common trait of these applicants was a great diversity of research and teaching qualifications, involving different disciplines; we were surprised by the informants' solid and comprehensive teaching backgrounds. Each of them had taught regularly at university level for more than ten years, and all were passionate about working with students.

As 18 of the 34 applicants were rejected in this round, we began wondering whether there might be a puzzle in the puzzle. Therefore, the initial phase of the study was followed by a comprehensive study of available documents related to the assessment process. By delving qualitatively into this complex issue, we aimed to provide insights into the risk of unintended consequences arising from the urge to standardize the assessment of teaching excellence (Subbaye & Vithal, 2017; Vardi & Quin, 2011). These insights may be helpful at governmental and institutional decision-making levels.

The interviewees were recruited using the snowball sampling method (Sadler et al., 2010). All respondents came from different departments and had no professional ties; however, they were familiarized with one another to some extent

during the digital preparatory sessions. We first identified two rejected applicants, and after that, we used respondents' knowledge of other applicants to recruit more rejected peers. All interviews were transcribed and we conducted a thematic analysis (TA), which is a widely used method in qualitative research 'for identifying, analyzing, and reporting patterns (themes) within data' (Braun & Clarke, 2006, p. 79). The inductive aspects of TA are particularly advantageous in new research areas where exploratory approaches are needed. In this study, we did not operate from any preconceived, interpretive framework; rather, we carried out a data-driven analysis (Braun & Clarke, 2006, p. 83). The objective was to identify patterns and theorize deeper meanings that emerged from the data, in addition to further contextualizing the themes according to existing pedagogical literature. Data saturation was evident by the third interview, and the last two interviews only brought minor new issues to the fore.

The open-ended interviews were conducted via the online platform Zoom and lasted between one and two hours. The informants were invited to talk freely for 10–15 minutes, without any interruptions, about their experiences with the application process. This session was followed by open questions in which the informants were asked to explain or further reflect on their experiences. The sound files were digitally transcribed and then read, reread, and coded as part of the TA. In this back-and-forth process, we regularly checked and compared the connections between the various pieces of text before developing a holistic presentation of the inherent meanings in the material. In TA, the goal is not to look for any objective or subjective truth or findings but to develop well-reflected new insights that may provide meaningful answers to the questions posed early in the study (Braun & Clarke, 2019, p. 594). The theme concept can be understood either as a summary of a concept or entity, a basic idea, or a meaning that facilitates an abstract and conceptual understanding of a particular matter (Braun & Clarke, 2023, p. 2).

The written data of the study also included close readings of the application guidelines, published on the university's homepage. The guidelines featured four assessment criteria, the applications' format, structure requirements, online supervision seminars, and the peer review process. The assessment criteria were (1) a focus on student learning, (2) professional development as a teacher practitioner, (3) an inquiry-based approach to teaching, and (4) a collegial attitude and practice (University, personal communication, 2022). In addition to the application guidelines and the transcribed interviews with informants, we also carried out close readings of the informants' applications, which were approximately ten pages each. Finally, we compared and analyzed the feedback reports that the informants had received from the evaluation committees.

Limitations of the study

This is a qualitative study based on data from a limited number of informants and documentary evidence from one Nordic higher education institution. This is not a satisfaction study or a best practice study, and we do not claim to generalize the results beyond the targeted sample. Rather, our aim is to contribute to a principled thinking of the theoretical foundations of the emerging field of educational merit awards that hopefully will stimulate considerations on educational merit schemes.

Results

We were surprised by the degree of consensus among the rejected applicants that surfaced from the coded data. We were also surprised to find that the first theme was highly psychological in nature. It concerns the informants' experiences of mental confusion and emotional shock due to the unexpected feedback presented by the evaluation committees. The next two themes, which we have called reflective logic versus "right-or-wrong" logic and the lack of constructive alignment respectively, revealed inconsistencies and weaknesses in the assessment design, which made the process difficult for the applicants to understand. These three themes are interrelated, and they are presented in the order in which they emerged during the analysis. In the following section, we discuss the extent to which indications of generic academic assessment principles are indicated or not indicated by the three themes and their coupling to the evolving neoliberalist tradition of assessment.

First theme: Mental confusion and emotional shock

The data revealed that the mandatory Zoom interview with the peer evaluation committee, comprising three persons from a diversity of disciplines, seemed to have a profound, negative impact on some informants. The individual interviews were conducted approximately one month after the submission deadline. All applicants who had fulfilled the formal application requirements were asked to individually meet their peer evaluation committee online. The data indicated that the interview served as a mental tipping point, during which the informants unanimously had experienced forms of mental confusion and emotional shock. Prior to this, they had not received any feedback on the expected outcome of the application. One informant (Respondent 1) described the experience as follows:

The interview was one of the most humiliating experiences of my life, so much so that I had suicidal feelings. I was about to do some self-harm to myself. I felt like cards were stacked against me right from the

start. I felt very discriminated against because there was no level playing field. By the end of the interview, it felt like I was being interrogated and that I had to fend off accusations. I do not think that I was allowed to justify my arguments at all. I froze at all the unpleasantness: 'We went through your application, and we don't feel you are formally qualified for this. Can you justify?' No other greeting, nothing! And I was in shock. (R1)

Other informants had assumed that the interview would be a dialogical conversation. However, sitting before the screen in front of a three-person committee, the experience was different. One informant said that he was put off because the interview was so negative and that he felt rejected even before it started. Others said they did not understand what the committee was after and became sleepless the following night: 'I didn't understand what they wanted to convey to me' (R4).

Before the application submission and the subsequent interviews, all informants had considered themselves devoted, innovative, and well-updated educators within their respective disciplines. However, six months after the interviews, some informants still experienced strong physical reactions. They identified emotions such as disappointment, anger, and frustration. Respondents experienced suicidal thoughts, tears, and feelings of humiliation to an extent that became life-disrupting. In one sense, such emotional reactions can be considered overreactions. However, the applicants were mature, empathic, and experienced educators, trained in assessing and handling emotional reactions among students; this context lends credence to the validity of their reactions. Their feelings of unease may be partly explained by a sudden, unexpected breach of expectations regarding the committee process and a loss of control in the interview situation. The following quote illustrates how one informant gradually became more uncomfortable when they realized that the rejection decision had likely been made before the interview:

Early in the interview, I had the impression that the committee had decided on a conclusion and that they were not open to changing their decision. I attempted to explain what efforts I had made, but they were mostly focused on my lack of self-reflection around these efforts. But that I didn't understand before the interview was finished. (R4)

The informants did not know what to expect or what was expected of them during the interviews. They associated the encounter with being placed before a jury and forced to make qualified guesses, while only the jury had the answer key. In particular, they were shocked that they did not qualify for the ETP award. Consequently, they spiraled into uncertainty and anxiety, wondering why they had

been invited to the interview in the first place. Half a year after the interviews, the informants still could not fully understand what they ‘had done wrong,’ and they were thoroughly bothered by this puzzle. The descriptions given by several informants—namely that they entered a ‘freeze’ response, became very confused, or were unable to talk—may be signs of a traumatic experience or mental shock, overwhelming their ability to handle the interview situation.

Second theme: Reflective logic versus right-or-wrong logic

Following the analysis of the open-ended interviews and the identification of the first theme, we considered the other datasets to acquire more information. By analyzing the informants’ applications, the feedback reports from the evaluation committees, and the institutional application guidelines, the second theme surfaced, concerning the discrepancies between a reflective logic and a right-or-wrong logic. All the informants had submitted reflective, chronologically structured applications. Support for the reflective logic was found in the online guidelines: ‘The application must show a coordinated presentation of your own competence and development as an educator, focusing on development work and leadership of education’ (University, personal communication, 2022). It was also specified that applications should be written as a reflection paper; furthermore, the applicants were expected to ‘reflect on the criteria for pedagogical merit and present arguments to support your fulfilment of the criteria’ (University, personal communication, 2022). The informants’ chronologically structured applications typically focused on their development as educators over time. Here are two typical examples:

In this reflection note, I will present an account of my experiences as a teacher, providing explanations for the factors that I believe have contributed to my growth and development as a teacher over the course of my career thus far, with emphasis on the last five years. (R1)

To give an account of my development as a teacher, I must provide an autobiographical account, from my first teaching appointment [...] through [my most recent position]. They are different educational ecosystems, and each offers different challenges and opportunities in the teaching/learning process. (R2)

The ETP award applicants were also offered information webinars in which they received lessons on how to structure the applications. Despite this, severe confusion among applicants prevailed: ‘It was a failure of communication in my opinion. It was not clear to me what I would be assessed on’ (R3). Yet another informant explained his choice of a chronological, reflective structure as follows:

I had a very distinct impression from all the informative seminars that the most important factor was the documentation of one's evolution as a teacher over a period of time. I had not even thought of organizing my application according to the four criteria. (R1)

However, close reads of the evaluation committees' feedback reports to the interviewed informants indicated that the committees' collective interpretation of the online guidelines differed considerably from the informants' interpretations. There was a mismatch between the online guidelines and the webinars on the one hand and the evaluation committees' perceptions of the guidelines on the other hand. While the informants had written chronological reflective papers, the committees' critical feedback was structured and standardized in a manner that did not make sense to the informants. They were surprised that the committees' feedback was not structured around the four criteria listed in the guidelines but rather around two subpoints connected to each criterion—a total of eight items (two subpoints for each criterion). In the online guidelines, the two subpoints of the first criterion, 'a focus on student learning,' were outlined as follows:

This means that the applicant

- has worked systematically over time to develop teaching that improves students' learning outcomes from the perspective of both the course and the program, and the applicant reflects on this development
- works systematically with and applies feedback from students and colleagues to develop teaching and the learning environment in courses and study program. (University, personal communication, 2022)

However, according to the guidelines, structuring based on the subpoints of the criteria was not an absolute requirement. They could easily be understood as examples or illustrations of potential competencies within the interpretation space of each criterion, and it would make little sense if the main criteria and their subpoints were to carry identical meanings. If so, the entire challenge would be akin to strictly adhering to a given recipe to produce a uniform result. Paradoxically, the assessors' feedback reports indicated that any alternative interpretation of the main criteria would automatically lead to rejection.

The conceptual confusion regarding the 'criteria' is thought-provoking. Several informants reported that they became completely confused when the first criterion as well as its two subpoints were referred to as 'criteria' in the feedback reports. While the informants were convinced that the 8,000-word reflective paper would work best if they followed a reflective, chronological logic, the evaluation committees seemed to follow a right-or-wrong logic structured strictly around the subpoints of the four main criteria. All five evaluation reports that we were allowed

to access followed the same standardized assessment template: the evaluation committees expected the applications to adhere to each subpoint rather than the main criteria. The assessment report for R1 cited below refers directly to the subpoints of the first main criterion:

[The first] assessment criterion reflects how the applicant works to place student learning at the center and how this is reflected in the applicant's view of learning in the subpoints, this is described in greater depth as meaning that the applicant 'has worked systematically over time with the development of teaching to improve students' learning outcomes from the perspective of both the course and the program, and they reflect on this development. In addition, the applicant is expected to have worked 'systematically with feedback from students and colleagues and applies this for further development of teaching and learning environments in courses and and programmes of study.' (Personal communication, 2022)

Our data indicates that instructions to the referees to focus exclusively on the subpoints placed strong limitations on the assessment process. The committee assessments thus became a test of adherence to an enforced interpretation of the main criteria rather than an opportunity for the applicants to demonstrate their own authentic competence within the much larger interpretation space associated with the main criteria.

At this point in the study, in search of an explanation for the highly perplexing mismatch that surfaced from the data material, we consulted a couple of invited peer reviewers. It appeared that the cohort of peer reviewers had been instructed, in a separate series of webinars, to ensure that specified "standards" were upheld. They were instructed that both subpoints of the four main criteria had to be filled in for an applicant to be accepted for the ETP award. While the informants followed a reflective logic, the peer reviewers had been instructed to apply a right-or-wrong logic in accordance with the subpoints. Consequently, a mismatch occurred between the two parties' interpretations of the criteria structure.

Third theme: The lack of constructive alignment

The third theme, the lack of constructive alignment, is closely related to both of the previous themes. The lack of constructive alignment represents the polarity of constructive alignment, as developed by Biggs (1996). This may for instance mean that written guidelines and assessment practices may be ambiguous or uncalibrated. In this study, the analysis of institutional guidelines, the informants' applications, and the committees' feedback reports demonstrated a lack of alignment regarding

the information about the criteria/subpoints between the guidelines/webinars and the committees' feedback reports. Indications of this mismatch is seen in this quote:

This is an example of horrendous design and evaluation. This is a checkbox! It needs an evaluation that takes the candidates seriously, that is open to different styles of development. The problem is that it is based on a closed evaluation system. But it is not appropriate to evaluate career paths because those are open, and there are possibilities and other aspects that may not have been considered. They are not evaluating individual merit; they are evaluating the merit of fitting into a certain structure. So, they haven't evaluated us as educators. (R2)

The following quotes further illustrate the informants' surprise at the proceedings: R1 said, 'I felt like it was a hit or miss and nowhere in between. They had no clue about what I was doing from [a professional] perspective,' while R5 said, 'I also had a feeling that it was a bit difficult for the committee to understand the complexity and the challenges that we have within our specific educational program.' Furthermore, there was a mismatch between the informants' and the evaluation committees' interpretations of information related to the application process. This mismatch seems to be caused by diametrically different views on the role of the main criteria and their subpoints. While the informants sought to make 'a coordinated presentation of [...] own competence and development as an educator, focusing on development work and leadership of education' as suggested in the application guidelines, the evaluation committees applied the criteria subpoints as the operative criteria, without informing the applicants (University, personal communication, 2022).

A criterion outlines what is to be assessed; however, there are clear distinctions between the criteria and their interpretations. Rather than representing feasible interpretations, the subpoints served as a mandated recipe at the expense of any other valid interpretation. This practice stands in opposition to statements encouraging flexibility in the application structure, such as 'it is up to the applicant to choose experiences and reflections that support the criteria for educational merit' and 'all experiences that have developed the applicant's pedagogical competence may provide a basis for recognition of merit' (University, personal communication, 2022). To address the confusion, the online seminar manager was consulted to authorize a valid interpretation of the criteria structure:

In a criterion description that describes something that can be achieved in more than one way, there is a need for an in-depth description. A challenge in interpretation is whether all subpoints are 'must' requirements, 'should' requirements, or [...] something that says that

an understanding of what is meant and how the criteria are fulfilled takes place in the peer-review processes. (Seminar manager, personal communication, September 23, 2023)

This confirms that ambiguities in the criteria structure remained unaddressed by the managers in charge of the ETP scheme. Issues of interpretation were left to the consultants who instructed the evaluation committees, a decision that was of no help to educators during the application process. However, repeated close reads of the committees' evaluation reports indicated that the peer reviewers were instructed to use principles associated with competency-based assessment (Gonczi, 1994) by reducing complex knowledge and skills to discrete and measurable competencies. This approach draws on an objectivist perspective of knowledge, requiring the fulfillment of the complete set of criteria and subcriteria to qualify for a pass (Bushway et al., 2018, p. 58); however, this requirement was never mentioned in the application guidelines.

Thus, the incompatibility of the assessment design caused the informants and the evaluation committees to adopt opposing approaches. On the one hand, the four criteria from the guidelines served as an invitation to write a reflective essay about the informants' development as teachers. On the other hand, the two subpoints of each criterion narrowed the interpretative options. The dilemma arose when the subpoints were treated as absolutes and referred to as the 'criteria' by the consultants and committees. This unwarranted interpretation demonstrated a lack of constructive alignment in the assessment design, leading to confusion:

They have given us an instruction manual based on a one-size-fits-all philosophy, and I don't approve of that. Take away these bullet points and let there be a scope of interpretation. (R1)

They [the reviewers] made a comment on not having followed a particular structure: However, that structure was never required; we were not required to follow that structure. Instead of using it as a guideline, they used it as a procedure. (R2)

While the university failed to apply the principle of constructive alignment in its assessment design, the applicants were candidly criticized by the peer reviewers for the same mistake: 'Some reference to constructive alignment (CA) has been made, but there are only reflections on alignment of content between courses, which is not a central aspect of CA' (R2 assessment report, personal communication, November 2022).

Summary and discussion

This study provides new insights into key aspects of the examined ETP system at a Nordic institution for higher education. As one of the very few studies focusing on the experiences of rejected applicants, this article traces assessment issues that seriously undermine trust in the ETP process. While the first theme identified in the empirical data, namely mental confusion and emotional shock, may initially seem to be a disproportionate reaction, further analysis uncovered that contradictory information had caused severe confusion and misconceptions among applicants. Consequently, applicants experienced adverse physical and emotional reactions, but were unable to conceptualize and theorize any deeper understanding of their own discomfort.

The analysis found that the applicants were surprised by a series of inconsistencies and contradictory statements in the assessment design—themes that we conceptualized as reflective logic versus right-or-wrong logic and the lack of constructive alignment, respectively. Based on our analysis, we argue that the mismatch between the applicants' expectations and the evaluation process in which they were enrolled can be understood on the basis of the disciplinary power of neoliberal governmentality (Raaper, 2017).

The inconsistent use of terminology, the ambiguous guidelines, and the opaque applicant interviews resemble previous studies on the evolving neoliberalism traits of Western universities. For instance, in a study of assessment policy discourses in two European universities, Raaper (2017) found that the assessment policy at the University of Glasgow, Scotland, was related not only to institutional contexts but also to 'globally dominant neoliberal discourses.' She further explained how issues of standardization and diffuseness are widespread features of neoliberal governmentality. Neoliberalism represents managerialism and marketization of higher education and comprises the use of performance measurement and delegation of management authority to new governance structures (Tolofari, 2005). Performance specifications are tight so as to detach policy formulation from policy execution. Typically, a strategic core of managers dedicates themselves to policy making while a more peripheral group of managers implement adopted policies. Respective obligations between principals and agents are spelled out in writing.

In the current study, ambiguities in the application guidelines led applicants astray in the application process. The problem was not the main criteria themselves, which were actually nonreferenced SoTL criteria. Rather, the problem faced by the applicants involved the subpoints added by local administrative staff, which in effect closed the open interpretation space of the main criteria. Thus, academically valid interpretations would inevitably lead to rejection because they did not comply with the administration-added subpoints. The intended academic reflection process

was thus transformed into a compliance exercise by replacing the main criteria with a standardized response template. The scheme failed to adhere to basic assessment principles, such as transparency, validity, reliability, authenticity and fairness.

Those principles imply that the assessment instrument should be well-grounded and justifiable, and sufficient to obtain the best possible understanding of the candidate's capabilities. The scrutinized ETP scheme was thus insufficiently constructed, compromising essential principles of assessment (Casey et al., 1997). The system adhered not to established principles of constructive alignment (Biggs, 1996) but rather to principles of competency-based assessment (Gonczi, 1994). This right-or-wrong logic undermined opportunities to demonstrate the depth and breadth of teaching competence. However, according to Royce Sadler, a leading higher education scholar, 'many—perhaps most—educational outcomes cannot be assessed as dichotomous states, although the competency assessment movement predominantly adopts that perspective' (Sadler, 2003, p. 4).

When both the validity and reliability of an evaluation are reduced, the principle of fairness comes into play. The applied report template illustrates this simplistic practice: 'Based on this assessment, the peers conclude that [the candidate] does not fulfill the criterion' (R1 assessment report, personal communication, 2022). The desire for adopting easily manageable assessment routines is understandable, but does not justify such simplistic solutions:

Appraisal processes all seek to capture some insight into how the teaching matches up to an ideal. This poses a difficulty—there is generally an assumption of some unique ideal way of teaching. Of course, this is not so. Teaching is, inevitably, context-dependent and even within a broad context, say laboratory work, or clinical studies, or tutorials, there will be different traditions, goals, ideals and exemplars from discipline to discipline (Casey et al., 1997, p. 470).

According to Saunders & Blanco Ramirez (2017), the measurement of a construct such as 'teaching excellence' is at constant risk of being oversimplified. Raaper (2017) noted that 'the government of assessment has become surrounded by discourses of administration that prescribe rules and replace traditional understanding of academic freedom with detailed authoritative directives' (p. 323). This study indicates that peers served on committees under the influence of a neoliberal policy of governance, uninformed by academic principles of academic assessment. Applicants tacitly expected to adhere to the latter, first and foremost transparency—a principle that includes all the necessary information about assessment methods, criteria structure, and how the final decision is derived, to avoid misunderstandings.

The shift to neoliberal ideology in academia was originally motivated by commercial interests, with the aim being to attract students in a competitive market. Consequently, the collegial and independent model of assessment has been replaced by a greater emphasis on administrative governance. This shift often requires academics to comply with rules and regulations, typically with less space for independent professional judgment. This is seen in the application of standardized routines, effectiveness, monetized rewards, and an objectivist view of knowledge with few, if any, references to generic principles of assessment.

In this study, the neoliberal approach is also demonstrated by a marked power imbalance between applicants and assessors as ‘there is no right of appeal or right of reply to the peer reviewers’ assessment and conclusion’ (University, personal communication, 2022). Such principles tend to result in less professional autonomy and academic freedom, and serve efficiency rather than learning.

Concluding remarks

This study found that the underlying, but not proclaimed, policy keywords for the evolving merit system appeared to be standardization, efficiency, competition, and the use of monetized incentives. While inconsistencies and minor deficiencies in a merit design are understandable in the early stages, this study indicates that in this case, a robust and conceptually anchored assessment framework is missing. Generic assessment principles may offer guidance during the design process as well as in the evaluation and revision of such evaluation instruments. Some may wonder why the institution embraces neoliberal principles at the expense of academic values and policies. Points to consider are the emergent rhetoric of quality, the growth in administrative positions, and the quest for time-efficient measures. The absence of an assessment policy may also constitute a conceptual vacuum paving the way for principles associated with neoliberal ideology.

Unique response requirements enable applicant comparisons; however, criterion-based assessment practices may operate on a different logic. Applications can be of equal quality while taking different approaches to valid interpretations of set criteria. In our globalized learning communities, diversity in teaching backgrounds alone has the potential to enrich practices far beyond formula-based expectations across disciplines. A successful future for the examined educational merit scheme would be an institutional, conceptual shift from teaching to learning. In this event, diverse applications could qualify for educational merit, steered not by compliance with prescribed teaching behaviors, but by institutional acknowledgement of reflective, original contributions—thereby embracing the pedagogical diversity needed to enhance future quality development of higher education.

The shift from traditional to neoliberal universities has caused extensive political and structural changes; however, some of those involved in merit processes may not be aware of the link between ideology and practice. A major task for future research would be to examine how the university community can recognize and address challenges posed by neoliberalism and managerialism, notably academic freedom and professional autonomy. The structural execution of power associated with managerialist approaches makes it imperative to provide counter-hegemonic discourses to understand and challenge the values and priorities of current practices.

Author biographies

Vidar Gynnild is a professor emeritus of higher education at Norwegian University of Science and Technology (NTNU, Trondheim). He has published widely on teaching, learning and assessment, and has been awarded distinctions for excellent educational development projects. Gynnild has also contributed at a vast number of seminars, workshops and conferences nationally and internationally.

Astrid Gynnild is a professor of Media Studies at the University of Bergen, Norway. Her current research focuses on technological innovation in the news media and on pedagogical innovation in higher education. She was awarded the title Excellent Teaching Practitioner (ETP) in 2019, as one of the first professors at the University of Bergen.

References

- Aurvoll, C., Larsen Damsgaard, H., and Enqvist-Jensen, C. (2023). Merit in the melting pot – The extent and significance of the Norwegian merit scheme. *Uniped*, 46(3): 201–212. <https://doi.org/10.18261/uniped.46.3.6>
- Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education*, 32(3): 347–364. <https://doi.org/10.1007/bf00138871>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2): 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Braun, V., & Clarke, V. (2019). Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health*, 11(4): 589–597. <https://doi.org/10.1080/2159676X.2019.1628806>
- Braun, V., & Clarke, V. (2023). Toward good practice in thematic analysis: Avoiding common problems and be(com)ing a knowing researcher. *International Journal of Transgender Health*, 24(1): 1–6. <https://doi.org/10.1080/26895269.2022.2129597>
- Bushway, D. J., Dodge, L., & Long, C. S. (2018). *A leader's guide to competency-based education: From inception to implementation*. Routledge. <https://doi.org/10.4324/9781003442844>
- Casey, R. J., Gentile, P., & Bigger, S. W. (1997). Teaching appraisal in higher education: An Australian perspective. *Higher Education*, 34(4): 459–482. <https://doi.org/10.1023/A:1003042830109>
- Chalmers, D. (2011). Progress and challenges to the recognition and reward of the Scholarship of Teaching in higher education. *Higher Education Research & Development*, 30(1): 25–38. <https://doi.org/10.1080/07294360.2011.536970>
- Felten, P. (2013). Principles of good practice in SoTL. *Teaching and Learning Inquiry*, 1(1): 121–125. <https://doi.org/10.20343/teachlearninqu.1.1.121>
- Førland, O., & Roxå, T. (2024). Establishing reward systems for excellence in teaching – The experience of academics pioneering a reward system. *European Journal of Higher Education*, 14(3): 451–469. <https://doi.org/10.1080/21568235.2023.2214713>
- Gonczi, A. (1994). Competency based assessment in the professions in Australia. *Assessment in Education: Principles, Policy & Practice*, 1(1): 27–44. <https://doi.org/10.1080/0969594940010103>
- Gourlay, L., & Stevenson, J. (2017). Teaching excellence in higher education: Critical perspectives. *Teaching in Higher Education*, 22(4): 391–395. <https://doi.org/10.1080/13562517.2017.1304632>
- Gunn, V., & Fisk, A. (2013). *Considering teaching excellence in higher education: 2007-2013. A literature review since the CHERI report 2007*.

- Higher Education Academy. Retrieved from <https://core.ac.uk/download/pdf/16460312.pdf>
- Gynnild, V. (2017). *Which are the key principles of assessment? A case study of policy documents*. ICERI 2017 Proceedings. http://library.iated.org/publication_series/ICERI
- Hancock, P. (2022). Employee recognition programmes: An immanent critique. *Organization*, 31(2): 381–401. <https://doi.org/10.1177/13505084221098244>
- Huggett, K. N., Greenberg, R. B., Rao, D., Richards, B. Chauvin, S. W., Fulton, T. B., Kalishman, S., Littlefield, J., Perkowski, L., Robins, L., & Simpson, D. (2012). The design and utility of institutional teaching awards: A literature review. *Medical Teacher*, 34(11): 907–919. <https://doi.org/10.3109/0142159X.2012.731102>
- Hutchings, P. & Shulman, L. S. (1999). The scholarship of teaching: New elaborations, new developments. *Change: The Magazine of Higher Learning*, 31(5): 10–15. <https://doi.org/10.1080/00091389909604218>
- Ka Yuk Chan, C., & Chen, S. W. (2024). Conceptualisation of teaching excellence: An analysis of teaching excellence schemes. *Assessment & Evaluation in Higher Education*, 49(4): 485-498. <https://doi.org/10.1080/02602938.2023.2271188>
- Larsen Damsgaard, H., Aurvoll, C., & Larsgaard, A. K. (2024). *Medaljen har også en bakside. Høyere utdanning sett innenfra – vitenskapelig ansattes erfaringer og refleksjoner* [The medal also has a reverse. Higher education seen from the inside – academic staff's experiences and reflections]. Universitetet i Sørøst-Norge.
- Lindberg-Sand, Å., & Olsson, T. (2008). Sustainable assessment?: Critical features of the assessment process in a modularised engineering programme. *International Journal of Educational Research*, 47(3): 165–174. <https://doi.org/http://dx.doi.org/10.1016/j.ijer.2008.01.004>
- Mårtensson, K., Roxå, T., & Olsson, T. (2011). Developing a quality culture through the Scholarship of Teaching and Learning. *Higher Education Research and Development*, 30(1): 51–62. <https://doi.org/10.1080/07294360.2011.536972>
- Miller-Young, J., & Yeo, M. (2015). Conceptualizing and communicating SoTL: A framework for the field. *Teaching & Learning Inquiry*, 3(2): 37–53. <https://doi.org/10.2979/teachlearninginqu.3.2.37>
- Ministry of Education and Research. (2017). *Quality culture in higher education*. <https://www.regjeringen.no/en/dokumenter/meld.-st.-16-20162017/id2536007/>
- Olsson, T., Mårtensson, K., & Roxå, T. (2010). Pedagogical competence - A development perspective from Lund University. In Å. Ryegård, T. Olsson, &

- K. Apelgren (Eds.), *A Swedish Perspective on Pedagogical Competence*, pp. 121–131. Division of Teaching and Learning, Uppsala University.
- Olsson, T., & Roxå, T. (2013). Assessing and rewarding excellent academic teachers for the benefit of an organization. *European Journal of Higher Education* 3(1): 40–61. <https://doi.org/10.1080/21568235.2013.778041>
- Raaheim, A., Grepperud, G., Olsson, T., Winka, K., & Pasteur Stø, A. (2020). *Evaluering av NTNUs system for utdanningsfaglig merittering* [Evaluation of NTNU's educational merit scheme]. NTNU. <https://www.ntnu.no/documents/1296117562/1309796912/Evaluering+av+system+for+merittering.pdf/38b9a295-e067-9d84-c18c-48dd8d4fa151?t=1643203422822>
- Raaper, R. (2017). Tracing assessment policy discourses in neoliberalised higher education settings. *Journal of Education Policy*, 32(3): 322–339. <https://doi.org/10.1080/02680939.2016.1257160>
- Sadler, D. R. (2003). *How criteria-based assessment misses the point*. ETL Conference, Queensland College of Art, Griffith University, Australia.
- Sadler, G. R., Lee, H-C., Seung-Hwam, R. & Fullerton, J. (2010.) Recruitment of hard-to-reach population subgroups via adaptations of the snowball sampling strategy. *Nursing & Health Sciences*, 12(3): 369–374. <https://doi.org/10.1111/j.1442-2018.2010.00541.x>
- Sandvoll, R., Winka, K., & Allern, M. (2018). Merittering som vitenskapelig tilnærming til undervisning [Pedagogical qualification frameworks as incentives for scholarly processes in education]. *Uniped*, 41(3): 246–258. <https://doi.org/10.18261/issn.1893-8981-2018-03-06>
- Saunders, D. B., & Blanco Ramírez, G. (2017). Against ‘teaching excellence’: Ideology, commodification, and enabling the neoliberalization of postsecondary education. *Teaching in Higher Education*, 22(4): 396–407. <https://doi.org/10.1080/13562517.2017.1301913>
- Skelton, A. (2004). Understanding ‘teaching excellence’ in higher education: A critical evaluation of the National Teaching Fellowships Scheme. *Studies in Higher Education*, 29(4): 451–468. <https://doi.org/10.1080/0307507042000236362>
- Subbaye, R., & Vithal, R. (2017). Teaching criteria that matter in university academic promotions. *Assessment & Evaluation in Higher Education*, 42(1): 37–60. <https://doi.org/10.1080/02602938.2015.1082533>
- Tolofari, S. (2005). New public management and education. *Policy Futures in Education*, 3(1), 75–89. <https://doi.org/10.2304/pfie.2005.3.1.11>
- University of Cape Town. (2024). Assessment policy for UCT. Retrieved May 20, 2025, from https://uct.ac.za/sites/default/files/media/documents/Policy_Assessment_2024.pdf

- Vardi, I., & Quin, R. (2011). Promotion and the scholarship of teaching and learning. *Higher Education Research & Development*, 30(1): 39–49.
<https://doi.org/10.1080/07294360.2011.536971>
- Wilcox, K. (2020). Interrogating the discourses of ‘teaching excellence’ in higher education. *European Educational Research Journal*, 20(1): 42–58.
<https://doi.org/10.1177/1474904120944783>
- Winka, K.. (2017). *Kartläggning av pedagogiska meriteringsmodeller vid Sveriges högskolor och universitet* [Mapping of educational qualification models at Sweden's colleges and universities]. Göteborgs Universitet.
<https://gupea.ub.gu.se/handle/2077/53276>
- Wood, M., & Su, F. (2017). What makes an excellent lecturer? Academics’ perspectives on the discourse of ‘teaching excellence’ in higher education. *Teaching in Higher Education*, 22(4): 451–466.
<https://doi.org/10.1080/13562517.2017.1301911>