

A critical view of The National Student Survey as a quality indicator in Norwegian higher education

Anders Dechsling,¹ Stefan Sütterlin,² Ricardo G. Lugo³ & Lars Rune Halvorsen^{4*}

¹Faculty of Teacher Education and Languages, Østfold University College, Norway; ²Faculty of Health, Welfare and Organisation, Østfold University College, Norway/Division of Clinical Neuroscience, Oslo University Hospital, Norway; ³Department of Information Security and Communication Technology, Norwegian University of Science and Technology, Norway & ⁴Faculty of Health, Welfare and Organisation, Østfold University College, Norway

ABSTRACT

Student surveys are an integral part of quality assessment in the education sector and play a vital role in the justification of policies and decisions on governmental, institutional and individual levels. Each year in Norway a governmental agency for quality assurance in education conducts a national survey inviting all registered second year bachelor's and master's students to provide online feedback on their perceived study quality. We discuss the limits of the results' interpretability in the light of previous research criticising the validity of student surveys for the assessment of educational quality in general and discuss in more detail the limitations in the chosen Norwegian example. This article aims to increase the awareness of these challenges and stimulate a science-based development of alternative assessment forms of educational quality. The relationship between the educational sector's core activity and the survey's focus is discussed; suggestions for paths to improvement are made. We argue further that the nationwide assessment lays bare the conceptual deficits that may be of equally high importance for educational system evaluations in other countries.

Keywords: *assessment; education; self-report; student survey; study quality*

Received: April, 2020; Accepted: April, 2021; Published: October, 2021

Through several public investigations and government acts, quality assessments have become an essential factor of policy-making in higher education, on the governmental as well as institutional level. In Norway, the Norwegian Agency for Quality Assurance in Education (NOKUT), under the Ministry of Education and Research, audits the

*Correspondence: Lars Rune Halvorsen, e-mail: lars.r.halvorsen@hiof.no

© 2021 Anders Dechsling, Stefan Sütterlin, Ricardo G. Lugo & Lars Rune Halvorsen. This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by-nc/4.0/>), allowing third parties to copy and redistribute the material in any medium or format and to remix, transform, and build upon the material for any purpose, even commercially, provided the original work is properly cited and states its license.

Citation: Anders Dechsling, Stefan Sütterlin, Ricardo G. Lugo & Lars Rune Halvorsen. «A critical view of The National Student Survey as a quality indicator in Norwegian higher education» Nordisk tidsskrift for pedagogikk og kritikk, Vol. 7, 2021, pp. 362–375. <http://dx.doi.org/10.23865/nipk.v7.2306>

effectiveness of higher education institutions' quality assurance systems and procedures for teaching and learning. What is arguably the best-known tool in NOKUT's toolbox is the annual "Studiebarometeret" (SB), a national student survey sent every autumn to each of the more than 60,000 second year students at the bachelor's and master's (i.e., 2nd or 5th year for master's) level. The participation rate in 2019 was 49%, with 17% of all questionnaires remaining incomplete (NOKUT, 2020). University students are a common tool of quality assurance in many educational systems. The surveys differ amongst others as a result of the various educational system structures, populations, size, political and societal priorities, or access to resources.

In this article, the Norwegian national student survey "Studiebarometeret" will serve as a platform for discussion. The degree to which these arguments can be generalised to other countries' comparable attempts to assess study quality or student satisfaction must be answered individually. We believe that this nationwide, representative, and annually conducted survey may provide a useful reference point of judgment regarding other systems. While study quality assessment is a widely known challenge, the Norwegian national student survey may serve as a particularly useful subject of discussion (see NOKUT, 2019a for further description, illustrations and information about SB). Its accessibility in the English language, adopted by the national educational sector covering both private and public educational institutions, reaches a significant sample of the student population, and receives strong media attention on a national scale.

The SB assesses the students' perceptions of educational quality in their study programmes and serves higher education institutions, future applicants, current students, and other entities interested in higher education. NOKUT states that a purpose of the national student survey is "to strengthen the quality work in higher education and give useful information about educational quality" (NOKUT, 2019a). In addition, NOKUT states that the results of survey are supposed to provide useful information for future students' choice of study programmes and institutions, as future study applicants are one of the target groups of the survey. It is further stated that the SB aims to be "an important aid to spread knowledge about study quality" (NOKUT, 2019a). While NOKUT also clearly states at various occasions that the SB assesses "perceived study quality", the students' subjective perception of study quality is obviously considered valuable enough to be seen as an important criterion for study choice.

In our opinion, perceived study quality expressed by students should be considered an interesting and valuable source of information, but not interpreted as a measure contributing meaningfully to study *quality*. It should therefore not be used as a tool informing decision- and policy-making on institutional or even governmental levels with the goal to increase scores in this survey. We will review empirical evidence collected over decades that establishes and repeatedly corroborates the notion that student self-reports on study quality are methodologically invalid tools for the assessment of actual study quality due to their profound susceptibility towards multiple

confounding influences. We conclude that these surveys are merely assessments of student satisfaction, and that they should be interpreted as such. We provide examples where the Norwegian national student survey SB fails to inform about the caveats of self-reports and aim to provide constructive criticism stimulating further discussion beyond national borders.

Assessment of study quality

Assessing study quality has long been a methodological challenge resulting in the development of a variety of complementary tools with varying strengths and challenges. Student surveys have been an integral part of these assessments for some time, in addition to more objective markers such as employability (employment rates of previous graduates) and academic performance indicators of teaching staff and their qualifications. Additionally, outward markers of quality such as output quality (achieved competencies at graduation in nationally or internationally comparable exams), perceived institutional reputation, research performance indicators with the assumption of a close association between research and teaching quality, funding situation, and class size are also used. Each of these and other indicators can serve as potential proxies and an operationalisation of the abstract concept of study quality, which lacks a universal consensus definition. This lack of definition is supported by Aarstad et al.'s (2019) findings of deficiencies in several proxies for study quality across the sector.

We argue that the quality of institutional decisions and governmental policies is at serious risk wherever the operationalisation of study quality is based mainly on perceived study quality obtained by student surveys. The proportionally excessive media attention student surveys receive may in part result from the high visibility, subjective relevance, and face validity within the student population and their organised representations. From a more optimistic viewpoint, media attention has the benefit of highlighting students' perspectives, experiences and general well-being. This benefit stands on its own merits, as students' well-being and their ability to have a strong voice in educational politics is in itself important in democratic institutions.

Assessing the educational institution's core activity

While many agree that quality assessment in higher education should relate as directly as possible to learning outcomes and competencies, the exact definition and aim of the core activities, and their relationship to perceived study quality from student perspectives, remain undefined and unmentioned in information provided about the SB. This lack of clarity leaves the question regarding the actual aim of the SB survey open, and does not explain which conclusions can or cannot be extracted from the data obtained. NOKUT itself states that the SB is to "promote quality in higher education. NOKUT's mandate is also to contribute such that society can trust the

quality in Norwegian higher education. Studiebarometeret [SB] is an important tool to disseminate knowledge on the educational quality” (NOKUT, 2019a). To define the actual core activity of the country’s higher education system, the government’s NOU report of 2006: 19 provides some guidance. The report from the Norwegian government states that educational quality is achieved through “free thinking, the search for the truth, understanding and acknowledgement, knowledge as a public right, and the value of a critical public sector” (NOU, 2006: 19, p. 12). This NOU report also describes:

the freedom to explore and dissemination of knowledge is grounded in society’s demand for a common, evidence based knowledge. This knowledge base is dependent on having trust throughout society such that the development of knowledge and what is considered relevant is not influenced by special interests, whether political, economical, or religious. This trust relies that research and teaching are built on proven and science-based models that use the current and valid methods and data, and are open for critical insights and testing.

An inspection of the SB’s survey items, however, reveals no items addressing the aforementioned outcomes such as evidence-based teaching, critical thinking or personal development. The criterion of “a public need for a common evidence based knowledge” is as little visible within the SB’s assessment questions as information regarding whether students receive an education based on scientific approaches or are trained in its core issues such as critical thinking and the ability to understand and process scientific literature. The absence of an assessment of these aspects makes it difficult to see how the SB can contribute to public trust in the educational system, which is defined as a goal in the NOU report (NOU 2006: 19).

The core activity in academia is founded and based on principles or ideals as the search for truth, namely developing knowledge, freedom of mind, critical thinking, and seeking understanding and insight (NOU 2006: 19). The primary objective of natural science, and some social sciences (e.g., psychology), is to seek truth and knowledge within the branch’s limited field (Gibbons, 1994). The actors within these sciences should, to a greater extent than the professional study instructors, know and master their science’s classical works and offspring, as well as determine what kind of questions the discipline is trying to answer. In addition, these actors should always orient themselves to the newest research within their field (Halvorsen et al., 2018). When evaluating the quality of a study programme, it is therefore reasonable to assume that assessing how well the programme delivers with regard to the intended core activity should be a central aspect.

However, it is worth noticing that only six out of 75 questions in SB are relevant to the academic ideals relating to the core activities of the study programmes. SB does not ask questions regarding the disciplines’ distinctive characteristics, how well the programmes master their disciplines’ sources of knowledge, or how good they are at including their own research or making use of newer scientific—and

thus international—literature. Further, SB scarcely contains any relevant questions regarding whether the students actually acquired knowledge from their discipline, or to what degree they are made capable of critical thinking, independent and self-responsible learning or processing of the newest literature in their field.

As opposed to branches of science and their aforementioned core activity related to the scientific method and active production of research, professional studies could be argued to differ in their definition of core activity in higher education—with possible consequences for the concept of study quality and implications for its assessment. Such are represented mainly in the health and social study domains, and include social work, nursing, and teacher education as examples. Professional studies, as taught in higher education institutions, can be characterised as instruments to solve different challenges in the welfare state. That does not imply that professions do not seek a scientific “truth”. However, professional studies rely on a compound of knowledge from a range of disciplines and areas of knowledge (Grimen, 2008). It is the professional studies’ objectives and tasks that determine which category of knowledge and skills, and their practical applications, are considered most relevant (Grimen, 2008; Halvorsen et al., 2018). Still, SB does not contain questions regarding to what extent each study programme has assembled the most appropriate compound of knowledge to solve the profession’s tasks, or how prepared the future professional will be after completion of their studies. As the SB hardly relates to the core activities for academic professions, prospective applicants—in their search for the institution that is most suited to train them to become the most skilful biologist, physicist, physician, teacher, nurse, social worker, etc.—are not presented with the relevant answers to their question in SB data.

The notable shortage of items addressing educational core activities stands in contrast to questions that could be best described as infrastructure-related. Such items enquire whether the lectures were engaging or digital tools were used; they also ask how many self-reported hours were spent on studies. This combination of items bears the risk that resulting scores rating infrastructure or self-reported engagement in the studies will be confused with actual study quality and thus be mistaken for being related to the educational institution’s core activity. As a result, incentives may be given to invest in infrastructure, entertainment and other predictors of student satisfaction, whereas those related to the development of knowledge—skills, the application of the scientific method, the experience of mastery in the sense of being challenged, and the necessary abilities to overcome intellectual challenges as stated in the previously mentioned NOU report—are neglected.

Student satisfaction and perceived quality: Confounding variables

Higher education is affected by the societal and political need for measuring performance in order to justify the distribution of funds and political actions. These needs are commonly considered to result mostly from economic pressure (Hazelkorn,

2015; Langan & Harris, 2019). After an adoption of business values in higher education (Birnbau, 2000; Langan & Harris, 2019), the observed shift of the attentional focus has been described as an understanding of the students' role as consumers or "customers" of educational services (Langan & Harris, 2019; Molesworth et al., 2009). Consequently, this may have facilitated the increased use and influence of using student satisfaction ratings and a confusion of student satisfaction as being a central result of educational activities, despite the lack of consensus regarding how to measure educational quality (Hazelkorn, 2015).

Students' teaching and teacher evaluations are widely used proxies in student quality assessment (Holland, 2019) and have increasingly often been subject to debates in the pedagogic literature. To the best of our knowledge, no comprehensive and science-based investigation into the predictors of student evaluation scores in the Norwegian SB has been carried out. Here we elaborate some known variables that have been shown to affect student evaluations. According to the international literature, a body of research reaching back to 1980, including pathway models, shows that grading leniency is a significant predictor for more positive student evaluations of teaching quality (e.g., Braga et al., 2014; Carrel & West, 2010; Howard & Maxwell, 1980). In a comprehensive review, Brockx et al. (2011) estimated that grading leniency accounted for approximately ten percent of student evaluations' variance, indicating a need for further research on other relevant factors. The causality problem and the lack of experimental and longitudinal designs is a methodological obstacle in this research.

Student evaluations, such as SB, do not control for confounding variables, introducing systematic biases. Hence, there are reasons to question the validity of student evaluation tools (e.g., Braga et al., 2014; Brockx et al., 2011; Chavez & Mitchell, 2019; Hessler et al., 2018; Murray et al., 2020; Schiekirka et al., 2015) and consequently their relevance for the assessment of study quality. Self-reported data is known to be very prone to situational variables such as, for example, mood states (e.g., Cohen et al., 1988), order effects (Atmanspacher & Römer, 2012), and retrospective biases (e.g., Smallwood & O'Connor, 2011). Recent research by Hessler et al. (2018) show how situational variables manipulated results from student assessments. Situational states were actively manipulated prior to assessment of students' perceived teaching quality, and in this experimental design, the availability of chocolate cookies during evaluation was associated with considerable positive effects on students' perceived teaching quality and the rated quality of the course material when compared to a control group with identical conditions without cookies (Hessler et al., 2018). The resulting effect sizes (Cohen's *d*) of cookie administration on perceived teacher quality and course material were 0.68 and 0.51, respectively. Situational context variables may influence study quality particularly strongly since the context in which the assessment takes place is closely related to the context that is to be assessed and is reflected upon—giving rise to the well-established response biases self-reported assessment is known to be prone towards.

The study by Hessler et al. (2018) illustrates that student satisfaction surveys have the potential to provide incentives for changes in teaching practices and may suggest that there can also be negative consequences for unpopular teaching practices regardless of the study quality as defined in the institution's core activity. While the popularity of teaching practices does impact seemingly unrelated constructs, such as perceived study quality, so do teacher characteristics. As recent research indicates, being taught by a female instructor or an instructor with a non-Western minority background can lead to lower perceived study quality, regardless of the fact that the internet-based course was completely identical between experimental groups (Chávez & Mitchell, 2019). An analysis of an online rating and student evaluation shows that young males with no accent were rated higher than others, and when a course was difficult, the overall ratings were lower (Murray et al., 2020). In addition, the researchers did not find any evidence linking student evaluations of teaching and research performance, which indicates that academic factors were irrelevant (Murray et al., 2020). To go even further, weather phenomena such as rain or low temperatures had a negative effect on the perceived quality of teaching professors (Braga et al., 2014). On the other hand, teachers that are more enthusiastic and have a good reputation systematically receive higher ratings for teaching quality, whereas the most relevant factors seem to be determined by structures and processes rather than the actual content (Schiekirka et al., 2015).

The above-mentioned examples raise the question as to whether these well-known vulnerabilities may be actively applied by teaching staff in order to increase their ratings. The term "academic gaming" refers to a way of playing along with decisions from management in higher education institutions in order to optimise the effects that these may have on one's own interests. This includes teaching staff's ways to react to centrally organised evaluation procedures (Ese, 2019). Ese suggests that academic gaming strategies may play a considerable role in the optimisation of student evaluations. Based on interview data from academic teachers in Norwegian higher education, teachers reported arranging popular lectures with particularly resourceful (guest) lecturers around the evaluation dates, practiced "teaching to the test", lowered the academic standards for exams, and handed out lecture notes beforehand. This was done with the knowledge that it would have detrimental effects on students' attention. Ese's research highlights actions taken by teachers to increase evaluation ratings, despite the fact they knew it would have detrimental effects on study quality.

The resulting institutional scores obtained by the national student survey SB receive considerable attention in the national media, particularly in education-related outlets, and are included in strategic decision-making processes on the institutional level. Media coverage puts a particular emphasis on comparative outcomes, i.e. the identification of particularly poorly or well-functioning areas or institutions. SB results do thus reflect the competitive nature of the higher education sector's recruitment and financing pressures and affect internal strategies, marketing profiles, recruitment efforts, and even budgeting priorities.

Contradictions in the self-presentation of the Norwegian national student survey

The empirically well-established susceptibility of self-reports on perceived study quality in student populations results in very questionable reliability and validity. Moreover, it is challenged by the high “face-validity” and arguably also the signal character of demonstrated concern about the student populations’ opinion. The economic advantages of online surveys in combination with the positive public impression “to care about study quality” may add further to the dominance of student satisfaction surveys as a tool of study quality assessment and as its misinterpretation as a representative and valid measure of quality. Given the described biases and shortcomings of self-reported quality assessment, and the to be expected reactions from teaching staff motivated to maximise their evaluation scores, administering a national student survey should be accompanied by measures taken to reduce the misuse, misinterpretation or overestimation of the survey’s data and methods.

The presentation of the SB on its official web page reveals, however, some striking contradictions that further question the theoretical foundation and overall design of this survey and its anchoring within the core activities that higher education institutions should identify with. On their information about the SB, NOKUT explicitly encourages the user to use the collected data on “perceived study quality” to “compare study programs and study places” (NOKUT, 2019a) and encourages potential future students to include these data as a criterion for their choice of study programmes and study places (NOKUT, 2019a). A convenient online function allows for direct comparison of scores across all registered study programmes, study areas, and institutions. The aim of the data obtained is described as “to spread the knowledge on study quality” (note that the reference to *perceived* study quality is not consistently used). In sum, the SB, and thus NOKUT, advise prospective students to rely explicitly on the perceived study quality of current students to compare scores of different study programmes and institutions, which creates the impression that perceived study quality is a direct measure of study quality.

Indicating some awareness of the problem described, NOKUT provides information in a less visible part of the download section of their webpages. Here, the authors refer to “systematic differences between study programs and various study areas” and mention that comparisons should rather be done between the same or relative similar study programmes. NOKUT also mentions that “students’ experience of educational quality is only one source of information and should therefore not be seen as the full truth. Other perspectives, such as teachers’ views and register data also say something about educational quality” (NOKUT, 2019b). This information, however, is less visible and refers to the relevance of alternative data and information sources that are neither available to the reader nor presented together with SB results.

These subtle “disclaimers” in the download section are in contradiction with the web presence’s functionality and the prominently posted contradictory encouragement to

compare between study programmes. These “warnings” follow the encouragement to do exactly the opposite, and appear like a “fig leaf” to counter methodological criticism, indicating a certain level of awareness about the methodological limitations of the gathered data. We argue that the promotion of and simultaneous advising against the common practice of using the SB mirrors the unresolved conflict between low construct validity and thus low data credibility versus the high and generally positive publicity of student satisfaction surveys in the media. Our point of criticism is therefore not to assess students’ perceived study quality, but rather the problem that follows from the very limited openness surrounding the limitations of this method—namely the unreflective use of these data and relative and passive acceptance of the misinterpretation of student survey data as a measure of quality, contrary to better knowledge.

Lack of coherence between self-report items and study aim

The scientific method, as it is followed by the social sciences, is an empirical method of acquiring knowledge that uses observation and the application of rigorous scepticism about what is observed, due to biases and errors that can distort how observations are interpreted. The scientific method involves observing, formulating hypotheses based on previous observations and theories, data collection from relevant samples, and testing of the hypotheses with appropriate analyses that lead to an eventual refinement of these hypotheses based on the evidence, which helps support theory development (Shadish et al., 2002). As the SB survey assesses personal beliefs and experiences, it would thus follow the social science methodology. The hypo-deductive method allows us to make predictions from hypotheses. The SB does not claim to be a scale that is constructed using psychometric standards, nor does it state any a priori hypothesis; it is thus a descriptive study. The SB survey may also be exploratory or informal in nature, and this may be fully sufficient for the purpose of a descriptive “mapping”. Nevertheless, the SB is described as a tool to assess the students’ perceived study quality. This requires, in the absence of further conditions, an assumption about the relationship of single items and study quality. Without this association between single items and the overall construct of study quality they aim to resemble, the informative value of single items remains more than questionable. This is the case in a number of items and thus can easily lead to misinterpretations.

In one item, the SB asks for the estimated hours per week students use organised teaching (lectures, etc.) vs. personal effort (Norw.: *egeninnsats*, or self-initiated learning). The informative value and assumed relationship with teaching quality remains very unclear. Both a particularly motivating and engaging teaching style, as well as particularly bad teaching and frequent teacher absence, causes the need for compensatory self-initiated learning, which may result in identical scores, yet reflect the very opposite of study quality. While it seems to be intuitive to assume a positive relationship between study quality and a teacher’s engaging learning style, the 2019

SB survey revealed that the aggregated data of 18 bachelor programmes in nursing suggest a negative association between self-initiated learning and perceived learning outcomes in the course (Norw.: *læringsutbytte*; NOKUT, 2019b). This effect seems to suggest that self-initiated learning can be both a consequence of good teaching or a compensation for bad teaching, and hardly sheds light on the perceived study quality, even within the same subject when assessed across institutions.

Besides items with a rather axiomatic relationship with study quality, other items included in the SB show an obvious face validity but demand a level of knowledge and (work) experience that can hardly be expected from second year bachelor students, which make up the majority of responders in the annual survey. Here, students are asked to judge the study's relevance for work life, which might be reasonable to ask the second year master's students. However, the majority of respondents to the annual SB survey are second year bachelor's students, and they are still in the phase of their (mainly full-time) education where they have very limited and only short term practical placement experience. Thus they are arguably unable to judge the relevance of theoretical knowledge for later practical work and hence considerably increase the risk of a Dunning-Kruger effect. The Dunning-Kruger effect (Kruger & Dunning, 1999) describes the relationship between individual skill levels and the accuracy of one's self-assessment. The inappropriateness of expecting an accurate self-assessment of one's ability to judge has consequences for student survey items that are beyond the responders' experience. The effect implies that one cannot fairly expect students to self-assess their own ability to respond to all items and thus recognise questions to which they may yet be unqualified to provide an informed answer (Dunning et al., 2004). Dunning et al. (2004, p. 69) conclude based on a considerable body of research that students

seem largely unable to assess how well or poorly they have comprehended material they have just read. They also tend to be overconfident in newly learned skills [...]. We suggest that policymakers and other people who makes real-world assessments should be wary of self-assessments of skill, expertise, and knowledge, and should consider ways of repairing self-assessments that may be flawed.

This well-established deficit of academic self-assessment is not exclusive to student populations or the academic context as such, but becomes relevant where the self-perceived learning success serves as a surrogate for the assessment of teaching quality, context and structures and their resulting effect on learning outcomes. This renders the perceived study quality a highly subjective measure that cannot be contrasted with any objective criterion. It remains, due to its high susceptibility to cognitive biases and subjective projections, of very limited use. We argue that assessing constructs that are beyond the students' personal experience is an inappropriate demand and not in line with basic principles of questionnaire construction. Consequently, this questions the validity of items in study quality assessments that aim to determine students' perceived study quality in the context of practical relevance.

In sum, we argue that student surveys assessing perceived study quality with a potentially high impact for policy-makers' decision making are not relieved from the basic scientific requirement that the items that were formulated must have a credible relationship with the underlying construct (here study quality). We further argue that study quality should be defined as close to the higher educational institutions academic core activity as possible to reduce the probability of influencing biases. We also stress the relevance of the respondent's capability built upon relevant experience to provide an informed judgment.

As the SB states on its website (NOKUT, 2019b), very different study programmes should not be compared with each other. As it can be assumed that ensuring a high study quality is equally important for all study programmes, it can be argued that this statement acknowledges that the survey does not appropriately fit all audiences, which undermines the actual need of a national student survey that is identical for all studies. If student surveys would be designed for more homogenous groups of study programmes, these weaknesses could be addressed. Furthermore, subsequent research should be directed at a systematic investigation to explore the range of particular policy documents and media attention within this domain. It would be particularly interesting to identify the amount of pressure this attention has on decision makers at the organisational level in the institutions within higher education, as well as the subsequent pressure on the teaching- and research staff with regard to archive results that look good at SB and the subsequent perceived reputation that follows media attention.

Conclusion

We argue that the SB does not provide valid insight into the quality of the institutions' core activities and does not allow for the prediction of which institution produces the best professionals. Neither the item formulation, item selection nor the categories addressed allow for such a conclusion. Misleading and partially contradictory claims and uses should thus be avoided.

The importance given to the SB results in the national media, in combination with the institutions' easing of measures to bias results in their favour, increases the probability of detrimental effects on the academic decisions by teaching staff and administrative managerial decisions taken at the leadership level. On an individual level, the fixation on positive evaluations creates a conflict between evaluation results and educational achievements. The very nature of national student satisfaction surveys tempts the user to compare the results between institutions and courses. We argue that cross-sectional comparisons between institutions do not add any relevant information about the quality of the institutions core activity to build competency due to numerous confounding factors outlined above.

Students can be expected to judge some aspects of study quality, but not all of them. The concept of educational quality can be compared to a similarly abstract and only partially accessible concept of health. One's own health can only be partially judged.

Chronic and fatal bodily dysfunctions can sometimes not, or only at a very late stage, be perceived as such. A thorough medical assessment requires multiple methods, of which self-reported symptoms are not necessarily the most relevant aspect of medical decision-making and particularly irrelevant with regard to diagnoses that are beyond the patient's ability to perceive.

We consider the assessment of perceived study quality as potentially interesting, as far as the subjective perception is interpreted in the context of a multi-method approach of study quality assessment. Potential alternatives that should be included in every discussion of national student satisfaction data have been widely discussed in the relevant literature and should accompany the media's interpretation. None of these alternatives are flawless, suggesting a multi-method approach including measures of employability, retrospective assessment of job-relevant qualifications obtained during studies, teachers' qualifications, formal criteria related to academic demands such as course requirements, grading leniency, and numerous others. Only a methodologically sound approach following scientific principles should inform institutional and nationwide policies.

Author biography

Anders Dechsling is currently a doctoral research fellow at Østfold University College's Faculty of Teacher Education and Languages. He specializes in research on Autism Spectrum Disorders.

Stefan Sütterlin is currently professor in psychology at Østfold University College's Faculty of Health, Welfare and Organisation. He is interested in experimental psychological research and investigates primarily conditions of emotional and behavioural regulation.

Ricardo G. Lugo is currently a researcher at Norwegian University of Science and Technology Department of Information Security and Communication Technology. His research interests include experimental and applied psychology of regulatory processes, cognitions, and performance.

Lars Rune Halvorsen is currently associate professor and head of department at Østfold University College's Faculty of Health, Welfare and Organisation. His research interests include the history and philosophy of science, science communication and educational methods.

References

- Aarstad, A. K., Hamberg, S., Oppegård, C., Pedersen, L. F., Sinderud, M. B., & Ulvevadet, B. (2019). *NOKUT's utredninger og analyser: Fagmiljø og utdanningskvalitet* (Rapport nr: 7-2019). NOKUT. https://www.nokut.no/globalassets/nokut/rapporter/ua/2019/fagmiljo-og-utdanningskvalitet_7-2019.pdf

- Atmanspacher, H., & Römer, H. (2012). Order effects in sequential measurements of non-commuting psychological observables. *Journal of Mathematical Psychology*, 56(4), 274–280. <https://arxiv.org/abs/1201.4685>
- Birnbaum, R. (2000). The life cycle of academic management fads. *The Journal of Higher Education*, 71(1), 1–16. <https://doi.org/10.1080/00221546.2000.11780813>.
- 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>.
- Brockx, B., Spooren, P., & Mortelmans, D. (2011). Taking the grading leniency story to the edge. The influence of student, teacher, and course characteristics on student evaluations of teaching in higher education. *Education Assessment, Evaluation and Accountability*, 23, 289–306. <https://doi.org/10.1007/s11092-011-9126-2>
- 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. <https://doi.org/10.1086/653808>
- Chavez, K., & Mitchell, K. M. W. (2019). Exploring bias in student evaluations: Gender, race, and ethnicity. *PS: Political Science & Politics*, 53(2), 270–274. <https://doi.org/10.1017/S1049096519001744>
- Cohen, L. H., Towbes, L. C., & Flocco, R. (1988). Effects of induced mood on self-reported life events and perceived and received social support. *Journal of Personality and Social Psychology*, 55(4), 669–674. <https://doi.org/10.1037//0022-3514.55.4.669>
- Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed self-assessment: Implications for health, education, and the workplace. *Psychological Science in the Public Interest*, 5(3), 69–106. <https://doi.org/10.1111/j.1529-1006.2004.00018.x>
- Ese, J. (2019). Defending the university? Academic's reactions to managerialism in Norwegian higher education [Doctoral dissertation, Karlstad University]. Karlstad University Digital Archive. <http://kau.diva-portal.org/smash/record.jsf?pid=diva2%3A1294669&dsid=1933>
- Gibbons, M. (1994). *The new production of knowledge. The dynamics of science and research in contemporary societies*. Sage Publications.
- Grimen, H. (2008). Profesjon og kunnskap. In A. Molander, & L. I. Terum (Eds.), *Profesjonsstudier* (pp. 71–86). Universitetsforlaget.
- Halvorsen, L. R., Løkke, J. A., & Granmo, S. (2018). Studentaktiv forskning som virkemiddel for å utdanne profesjonsutøvere som også kan atferdsanalyse. *Norsk tidsskrift for atferdsanalyse*, 45, 87–100.
- Hazelkorn, E. (2015). *Rankings and the reshaping of higher education: The battle for world-class excellence*. Springer.
- 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>
- Holland, E. P. (2019). Making sense of module feedback: Accounting for individual behaviours in student evaluations of teaching. *Assessment & Evaluation in Higher Education*, 44(6), 961–972. <https://doi.org/10.1080/02602938.2018.1556777>
- Howard, G. S., & Maxwell, S. E. (1980). Correlation between student satisfaction and grades: A case of mistaken causation? *Journal of Educational Psychology*, 72(6), 810–820. <https://doi.org/10.1037/0022-0663.72.6.810>
- Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77(6), 1121–1134. <https://doi.org/10.1037/0022-3514.77.6.1121>
- Langan, A. M., & Harris, W. E. (2019). National student survey metrics: Where is the room for improvement? *Higher Education*, 78, 1075–1089. <https://doi.org/10.1007/s10734-019-00389-1>
- Molesworth, M., Nixon, E., & Scullion, R. (2009). Having, being and higher education: The marketisation of the university and the transformation of the student into consumer. *Teaching in Higher Education*, 14(3), 277–287. <https://doi.org/10.1080/13562510902898841>.
- Murray, D., Boothby, C., Zhao, H., Minik, V., Bérubé, N., Larivière, V., & Sugimoto, C. R. (2020). Exploring the personal and professional factors associated with student evaluations of tenure-track faculty. *PLoS ONE*, 15(6), e0233515. <https://doi.org/10.1371/journal.pone.0233515>
- NOKUT. (2019a). *Studiebarometeret*. <http://studiebarometeret.no/no/artikkel/2>
- NOKUT. (2019b). *Studiebarometeret*. <http://studiebarometeret.no/no/artikkel/5>
- NOKUT. (2020). *Studiebarometeret 2019: Dokumentasjonsrapport*. https://www.nokut.no/globalassets/studiebarometeret/2020/hoyere-utdanning/studiebarometeret-2019_dokumentasjonsrapport_2-2020.pdf

A critical view of The National Student Survey as a quality indicator in Norwegian

- NOU 2006: 19. (2006). *Academic freedom*. Ministry of Education and Research. <https://www.regjeringen.no/en/dokumenter/NOU-2006-nr-19-Academic-freedom/id392466/>
- Schiekirka, S, Feufel, M. A., Herrmann-Lingen, C., & Raupach, T. (2015). Evaluation in medical education: A topical review of target parameters, data collection tools and confounding factors. *German Medical Science, 13*. <https://doi.org/10.3205/000219>
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin Company.
- Smallwood, J., & O'Connor, R. C. (2011). Imprisoned by the past: Unhappy moods lead to a retrospective bias to mind wandering. *Cognition & Emotion, 25*(8), 1481–1490. <https://doi.org/10.1080/02699931.2010.545263>